TU WIEN Informatics

Dark Patterns for Good?

Exploring End-User Perspectives on Bright Patterns to Counteract Perceived Social Media Problems

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieurin

im Rahmen des Studiums

Media and Human-Centered Computing

eingereicht von

Doris Maria Rhomberg, BSc
Matrikelnummer 01649642

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Ass.-Prof. Dr.x Katta Spiel
Mitwirkung: DI Janis Lena Meißner, MRes PhD

Wien, 21. Oktober 2024

Doris Maria Rhomberg	Katta Spiel

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.at

TU WIEN Informatics

Dark Patterns for Good?

Exploring End-User Perspectives on Bright Patterns to Counteract Perceived Social Media Problems

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieurin

in

Media and Human-Centered Computing

by

Doris Maria Rhomberg, BSc
Registration Number 01649642

to the Faculty of Informatics

at the TU Wien

Advisor: Ass.-Prof. Dr.x Katta Spiel
Assistance: DI Janis Lena Meißner, MRes PhD

Vienna, 21st October, 2024

Doris Maria Rhomberg	Katta Spiel

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.at

Erklärung zur Verfassung der Arbeit

Doris Maria Rhomberg, BSc

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 21. Oktober 2024

Doris Maria Rhomberg

Danksagung

Ich möchte meiner Betreuerin Janis Lena Meißner und meinen Bezugspersonen in Amerika Hauke Sandhaus und Colin Gray, die mich auf meinem Weg begleiteten, meinen tiefsten Dank aussprechen. Ich bin sehr dankbar für ihre Unterstützung und Bestärkung, die mir geholfen haben, motiviert zu bleiben, sowie für das wertvolle Feedback, das sie mir während des gesamten Prozesses dieser Arbeit gaben.
Janis Lena Meißner war mir eine unglaubliche Hilfe bei der Suche nach einem geeigneten Thema, das nicht nur meinen Interessen, sondern auch meinen bisherigen Erfahrungen entspricht. Sie ermutigte mich, im Ausland zu forschen, stellte relevante Ressourcen zur Verfügung und machte mich mit Colin Gray bekannt, einem Experten im Thema Dark Pattern. Ohne ihre Unterstützung wäre diese besondere Chance und einmalige Erfahrung, an der ich beruflich und persönlich gewachsen bin, nicht möglich gewesen.
Hauke Sandhaus empfing mich herzlich in New York City und war mein Gastgeber an der Cornell Tech, wo ich einen Teil meiner Forschung durchführte. Hauke war nicht nur ein guter Freund, sondern auch ein ausgezeichneter Forschungspartner. Er ermutigte mich, als Co-Autor an einem CHI-Workshop-Paper mitzuschreiben, was eine großartige Erfahrung war.
Colin Gray hieß mich an der IU Bloomington, wo ich den letzten Teil meiner Forschungsarbeit abschloss, als Gastgeber willkommen. Mit Colins Hilfe konnte ich ein typisch amerikanisches Collegeleben kennenlernen und Einblicke in die Forschungsarbeit in den USA erhalten. Ich hatte die Ehre, an Lab-Meetings teilzunehmen und mich an Forschungsprojekten zu beteiligen.

Mein besonderer Dank gilt auch Katta Spiel, nim mein Interesse für Ethik in der Informatik geweckt hat, und mich in organisatorischen Belangen unterstützt hat.

Ich danke auch allen Teilnehmer:innen meiner Umfrage und Studie, die viel Zeit aufwendeten und ihre Gedanken und persönlichen Erfahrungen im Zusammenhang mit sozialen Medien teilten.

Abschließend möchte ich mich bei meinen Kommiliton:innen, Freund:innen und Familienmitgliedern für ihre Hilfe beim Korrekturlesen und ihre moralische Unterstützung bedanken, sowie bei all den wunderbaren Menschen, die ich im Ausland kennenlernte, die mich inspirierten und einen langanhaltenden Eindruck hinterließen.

Acknowledgements

I want to express my deepest gratitude to my supervisor, Janis Lena Meißner, and collaborators, Hauke Sandhaus and Colin Gray, who have all followed along my journey. I am thankful for their support and encouragement, helping me stay motivated, as well as the thought-provoking and valuable feedback they provided throughout the thesis process.

Janis Lena Meißner has been a wonderful help in finding a research topic that not only aligns with my interests but also builds on my previous experiences. She also encouraged me to pursue research abroad, provided relevant resources, and introduced me to Colin Gray, an expert in dark patterns. This incredible opportunity and once-in-a-lifetime experience, where I grew professionally and personally, would not have been possible without her support.

Hauke Sandhaus welcomed me to New York City and hosted me at Cornell Tech, where I conducted part of my research. Beyond being a great friend, Hauke was an exceptional research partner. He encouraged me to co-author a CHI workshop paper, which was a great experience.

Colin Gray welcomed and hosted me at IU Bloomington, where I completed the final part of my research. Colin offered me the wonderful opportunity to experience a typical American college environment and gain insights into how research is conducted in the US. I had the pleasure of attending lab meetings and participating in research projects.

My special thanks also go to Katta Spiel, who sparked my interest in ethics in computer science and supported me in organizational matters.

I am also thankful to all my survey and study participants for generously spending their time and sharing their thoughts and personal experiences related to social media.

Lastly, I want to extend my thanks to my classmates, friends, and family members for their editing assistance and moral support, as well as to all the wonderful people I met abroad, who have left a lasting impact and inspired me throughout this journey.

Kurzfassung

Dark Pattern sind Design Tricks, mit denen Nutzer:innen dahingehend manipuliert werden, Handlungen zu tätigen, die sie andernfalls möglicherweise nicht vornehmen würden. Sie sind weit verbreitet und unter anderem in Sozialen Medien zu finden, wo sie Nutzer:innen dazu verleiten, mehr Daten als beabsichtigt preiszugeben oder länger auf den Plattformen zu verweilen. Während Dark Patterns Nutzer:innen zu den Zielen eines Unternehmens führen, lenken Bright Patterns sie hin zu Handlungen, die mit ihren persönlichen Zielen übereinstimmen. Spontan empfundene Wünsche stehen dabei nicht im Fokus. Frühere Studien beschäftigten sich bereits mit Bright Pattern in Cookie-Zustimmungsanfragen und im E-Commerce. Deren Auswirkungen im Kontext Sozialer Medien wurden jedoch bisher nicht untersucht.

Diese Masterarbeit untersucht aus der Perspektive der Endnutzer:innen, wie Bright Pattern als mögliche Strategie eingesetzt werden können, um wahrgenommene Probleme in sozialen Medien zu mildern. Dabei werden bewährte Konzepte von Dark Pattern zugunsten der Nutzer:innen umgekehrt. Die Arbeit umfasst eine Literaturrecherche über Taxonomien, eine Umfrage zur Messung der Wahrnehmung von Dark Pattern durch die Nutzer:innen, die Gestaltung von drei Bright Pattern Mockups sowie Nutzer:innenstudien zur Bewertung der neuen Ideen. Durch Prüfung und Kombination bestehender Dark Pattern Klassifizierungen wurde eine neue Taxonomie mit 15 Low-Level Dark Pattern entwickelt, die speziell für Soziale Medien gilt. Die Bewertung der Mockups dieser 15 Dark Pattern mittels eines angepassten User-Experience-Fragebogens zeigte eine Korrelation zwischen den Werten der Täuschung und der Attraktivität, was auf Folgendes hindeutet: Je täuschender ein Dark Pattern ist, desto unattraktiver ist es. Zudem wurden Dark Pattern als aufdringlich, nervig und unfreundlich empfunden, wobei das Pattern *Forced Access* von allen Pattern am schlechtesten bewertet wurde. Drei ausgewählte Dark Pattern (*Nagging*, *False Hierarchy* und *Toying with Emotion*) wurden in Bright Pattern umgewandelt, die darauf abzielen, die aufgewendete Zeit zu verringern, die Verbreitung von Fake News zu verlangsamen sowie sinnvollere und vielfältigere Inhalte zu präsentieren. Die durchgeführten Nutzer:innenstudien verglichen Bright Pattern mit den entsprechenden Dark Pattern und ergaben, dass alle Pattern zwar als nervig empfunden wurden, die Bright Pattern jedoch insgesamt besser abschnitten. Diese Arbeit legt nahe, dass Dark Pattern zum Wohle der Nutzer:innen umgewandelt werden können, auch wenn ethische Bedenken hinsichtlich des Einsatzes gezielter Manipulation bestehen bleiben. Sie trägt zur Literatur über Dark Patterns sowie Bright Patterns bei.

Abstract

Dark patterns are design tricks used to manipulate users into actions they might not have otherwise taken. They are prevalent on the internet and commonly found in social media, where they lead users to overshare their data or overspend their time on the platform. While dark patterns lead users to a business' objectives, bright patterns guide users toward actions that align with their personal goals without prioritizing their immediate desires. Previous studies explored bright patterns in areas like cookie consent requests and E-Commerce, but their impact in the social media context has not been studied.
This master's thesis investigates the use of bright patterns as a potential strategy to address perceived social media problems from an end-user perspective. The approach involves inverting proven dark pattern concepts to benefit users. This work involves a literature review of taxonomies, a survey to assess users' perceptions of dark patterns, the design of three bright pattern mockups, and user studies to assess the new ideas. By examining and combining existing dark pattern taxonomies, a new taxonomy of 15 low-level dark patterns specific to social media was developed. Surveying mockups of these 15 dark patterns, using an adapted user experience questionnaire, revealed a correlation between the deceptive and the attractive scores, suggesting that more deceptive dark patterns are less attractive. Additionally, dark patterns were perceived as pressuring, annoying, and unfriendly, with *Forced Access* rated the worst among all patterns. Three selected dark pattern types (*Nagging*, *False Hierarchy*, and *Toying with Emotion*) were inverted into bright patterns aimed at reducing time spent, decelerating the spread of fake news, and presenting more meaningful and diverse content. User studies comparing these bright patterns with their corresponding dark patterns found that while all patterns were perceived as a nuisance, bright patterns scored better overall. This thesis suggests that dark patterns can be repurposed for the user's good, though ethical concerns about targeted manipulation persist. It contributes to the literature on dark patterns as well as bright patterns.

Contents

CHAPTER 1

Introduction

Dark patterns are design practices employed to influence and guide users toward actions they may not have taken otherwise [14]. In recent years, there has been a strong interest among Human-Computer Interaction (HCI) researchers in studying these malicious design techniques, evidenced by a significant increase in papers that have been published. In 2018, searching for dark patterns in the ACM Digital Library resulted in fewer than 10 papers, whereas by 2020, the number exceeded 30, indicating a three-fold increase in just two years [47]. As of August 2024, typing the term into the search bar yielded over 36,000 results[1].

This increase in interest is likely due to the pervasiveness of dark patterns, as they are all across user interfaces, particularly found in E-Commerce [52, 59], Gaming [4, 82], and Social Media [27, 56, 57]. In an automated analysis of 11k worldwide shopping websites, Mathur et al. identified 1,818 instances of dark patterns in 1,254 websites [52]. Kirkman et al. found 3,744 different dark patterns in 2,417 consent dialogs in another automated detection of dark patterns [42]. Also high is the pervasiveness of dark patterns in mobile apps. Di Geronimo et al. walked through 240 popular free Android apps belonging to eight different categories to detect malicious designs and found that 95% of the apps included at least one dark pattern [26].

This pervasiveness may be explained by some designers copying the design of others [26] or by pressure from the marketplace [21]. Another reason could be that companies simply know that dark patterns are effective in manipulating users [46]. Luguri and Strahilevitz assessed the efficiency of dark patterns in manipulating users and found out that a mild dark pattern doubled the rate of participants accepting a company's preferred option, and an aggressive dark pattern almost quadrupled the rate of acceptance [46]. Similarly, Utz et al. tested if the number, design, or pre-selection of choices in a cookie consent

[1] `https://dl.acm.org/action/doSearch?AllField=dark+patterns`, last accessed: 2024-08-03

request impacts users' decisions. Their findings demonstrated that the more choices participants had, the more likely they were to decline cookies. Results also showed that highlighting an option increased acceptance rates by 29.9% on mobile devices and by 27.5% on desktop devices. The pre-selection of choices led to acceptance by three in ten mobile users and one in ten desktop users, while almost no users accepted the option when it required an opt-in choice [79]. In social media, dark patterns manipulate users to overshare data [27] or to get them hooked [56, 65]. Mathur et al. summarized that dark patterns achieve to reduce three different kinds of individual welfare: financial loss, invasion of privacy, and cognitive burden. Furthermore, they highlighted that, from an individual autonomy perspective, some dark patterns maximize user engagement and enable addiction by a choice architecture influencing users to make choices they would not have done otherwise, denying a user choice, hiding available choices, or burdening the exercise of choice [53].

1.1 Motivation and Aim of the Work

Given the pervasiveness and effectiveness of manipulating users with dark patterns, this thesis explores whether dark patterns can be repurposed for the benefit of users. While existing literature mainly focuses on combating dark patterns, including raising users' awareness and creating countermeasures, the goal of this thesis is to understand when and how persuasive design techniques can be used positively, drawing on insights from the effective strategies employed by dark patterns. The notion of repurposing dark patterns is supported by insights from existing literature: A participant in a different study mentioned that it is impossible to stop companies from utilizing them [50] and Gray et al. highlighted that persuasive strategies, including dark patterns, possess the potential to be employed for both positive and negative purposes [36].

Repurposing dark patterns for the users good leads to the concept of bright patterns. Researchers use this antonym to describe persuasive patterns that prioritize users' goals over their desires and companies' objectives [67]. The existing limited research on inverting dark patterns into bright patterns has shown that this switch could lead to a lower cookie consent acceptance rate [37] and higher perceived freedom of choice and transparency [78].

This thesis seeks to invert dark patterns into bright patterns and assesses how they are perceived by users. It concentrates on the context of social media, whereby bright patterns will be used to help users overcome three of their perceived social media problems. The aim of this thesis is not to counteract dark patterns or resolve the issues they create but rather to use bright patterns (inverted dark patterns techniques) to address social media problems. An end-user perspective is chosen since this, on the one hand, helps to determine if a pattern is a dark pattern or not [16], and on the other hand, to identify users' needs and subsequently imagine solutions like introducing friction to combat infinite scrolling [10]. This thesis explores a whole new area by investigating both dark and bright patterns from an end-user perspective in the context of social media.

1.2 Research Questions

To achieve this thesis' aim of exploring strategies for utilizing dark pattern concepts so that users derive a positive impact from them, the following three research questions were addressed, each representing a progressive step toward the overarching goal:

1. RQ1: What are known problematic design patterns that manipulate users on social media, and how do users perceive them?

The first part of this question is answered by a literature review, where various taxonomies of dark patterns were investigated to find known dark patterns that are proven to be manipulative. Dark pattern types were gathered, duplicates were removed, and similar ones were merged, to establish a new taxonomy of dark patterns prevalent on social media. Subsequently, the users' perception of the resulting dark pattern types was tested in a survey to answer the second part of this question. Understanding how users perceive these dark patterns helped in further investigation and in deciding on which dark patterns to concentrate on in this master's thesis.

2. RQ2: How could known dark pattern mechanics be repurposed as countermeasures against users' perceived social media problems?

This question is answered by an exploratory design approach, gathering inspiration from research and existing dark patterns. The aforementioned survey also aimed to uncover users' perceived social media problems to identify key areas requiring intervention. Based on the identified issues, new ideas for bright patterns designed to counteract them were elaborated.

3. RQ3: How do users perceive the newly designed bright patterns compared to the corresponding dark patterns?

User studies were conducted to determine if users like the newly created bright patterns to counteract social media problems. The assessment focused on users' perception and acceptance towards the patterns. This resulted in a comparative analysis of dark patterns and bright patterns within the realm of social media.

1.3 Methodological Approach

The presented work uses a mixed-method study, including qualitative and quantitative methods, to answer the aforementioned questions. The following four steps were performed:

1. **Literature Review**
   The initial stage includes an extensive review of the literature, investigating the definition and user research of both dark patterns and bright patterns. Furthermore, the different types of dark patterns employed on social media, along with their characteristics, are examined. Moreover, papers examining the distinction between manipulation and persuasion are presented.

2. **Survey**
   In the second step, a survey is conducted with three main goals: (1) assessing the previously collected dark patterns used on social media, (2) identifying users' perceived problems with social media, and (3) understanding users' social media usage and habits. The former is done by presenting users with the different types of dark patterns and asking them to fill out an adapted version of the User Experience Questionnaire (UEQ)[2]. The second objective is achieved by a combination of Likert-scale questions and open-ended questions, and the third goal is accomplished by single and multiple-choice questions.

3. **Design**
   The third step involves the exploration of bright patterns as a countermeasure against social media problems. Therefore, the previously acquired knowledge of steps one and two is combined within different design thinking methods to establish and design bright patterns. After conceptualizing the ideas, prototypes are built.

4. **Assessment**
   In the final step, bright patterns as well as their corresponding dark patterns are assessed, focusing on users' perception and acceptance. Finally, the results of a social media usage with dark patterns are compared to one involving bright patterns.

These steps allowed an exploration and investigation of utilizing dark patterns for the benefit of users by inverting them into bright patterns. While the former two steps focus on dark pattern research, the latter two center around the research on bright patterns. These bright patterns should foster change in user behavior in the context of social media. This thesis presents a new taxonomy of 15 low-level dark patterns found in social media, as well as how survey participants perceived them. Subsequently, three dark patterns were inverted into bright patterns aiming to reduce time spent, decelerate the spread of fake news, and present more meaningful and diverse content. Finally, the thesis compares the three selected dark pattern types with their corresponding bright patterns resulting from the user study data. Additionally, it presents ethical as well as design dilemmas that arise when creating bright patterns.

[2] `https://www.ueq-online.org/`, last accessed: 2024-08-07

1.4 Structure of the Thesis

This thesis is structured as follows: In Chapter 2, existing literature on dark as well as bright patterns is presented, focusing on relevant user studies and different taxonomies. Chapter 3 explains the research approach, giving details of each step done during this thesis. The subsequent two chapters present findings from the dark pattern research and the bright pattern research. Chapter 4 presents the findings from analyzing dark patterns, including the developed taxonomy, as well as the findings from the survey. Chapter 5 presents the findings from exploring bright patterns, including the newly designed bright patterns, and the findings of the assessment phase. All these findings are then discussed in Chapter 6, drawing references on existing literature, discussing the ethical and design dilemma behind inventing bright patterns, as well as contributions, limitations, and future work. In the end, Chapter 7 summarizes this thesis.

CHAPTER 2

Related Work

This section presents an overview of relevant literature on dark patterns as well as bright patterns, followed by a discourse on influential design practices.

2.1 Dark Patterns

Current literature involves a variety of different dark pattern definitions. The term was first coined in 2010 by Harry Brignull, who now refers to them as deceptive patterns. According to him, these are *"tricks used in websites and apps that make you do things that you didn't mean to, like buying or signing up for something"* [14]. In contrast, Gray et al. use this term to *"define instances where designers use their knowledge of human behavior (e.g., psychology) and the desires of end users to implement deceptive functionality that is not in the user's best interest"* [36]. Sánchez highlights two problems with the term dark patterns. On the one hand, it induces racist stereotypes, and on the other hand, it is too vaguely formulated, which leads to confusion among practitioners [75]. The two popular definitions of Brignull [14] and Gray et al. [36] illustrate this variety. While the former is more outcome-oriented, focusing on the misleading actions users are tricked into, the latter emphasizes the intentional misuse of human behavior knowledge to deceive users.

Mathur et al. say that dark patterns coerce, steer, and deceive users into making decisions that they would not make under complete information and when presented with alternatives [52], what partly overlaps with the techniques listed by Conti and Sobiesk like coercion, distraction, exploiting errors, forced work, obfuscating desired content, restricting or masking functionality, and deception or misrepresentation, among others [21]. Mathur et al. also outlined five different dimensions to determine whether something qualifies as a dark pattern or not: asymmetric, covert, deceptive, hides information, and restrictive. *Asymmetric* examines whether the design puts unequal weights or burdens on available choices, *covert* assesses if the effect of a user's choice is hidden, *deceptive* evaluates if an interface causes false beliefs, *hide information* checks

if crucial information is hidden or delayed from the user, and *restrictive* considers whether the interface limits the available choices [52]. Mildner et al. used these five characteristics to assess the malice of interfaces. The authors prompted users to evaluate their perceptions of these patterns using these dark pattern characteristics outlined by Mathur et al. [52]. They found a significant association between the characteristics asymmetric and restrictive with the malice score. However, the other three characteristics did not yield significant associations. Thus, they suggest that the maliciousness of an interface can be determined by considering the characteristics ratings, but the model needs to be adjusted for proper assessment [54].
In a different research, Mathur et al. came up with four different perspectives for analyzing dark patterns and their effects on individuals and society, each assessing how dark patterns modify the choice architecture. While the lens of the *Individual Welfare* judges how the choice architecture is adjusted at the expense of the individual welfare, the *Collective Welfare* has a broader perspective, assessing how dark patterns affect the collective welfare. The lens under the *Regulatory Objectives* describes how specific regulatory objectives are undermined due to dark patterns. Lastly, the *Individual Autonomy* lens considers dark patterns as undermining individual decision-making [53]. In order to decide if a particular interface is a dark pattern, the authors propose the question of whether it undermines individual welfare [53]. Similarly, Zagal et al. presented questions that can be asked in order to check if a dark pattern is present [82].

2.1.1 Background of Dark Patterns

The existing literature on dark patterns is extensive, ranging from the pervasiveness of dark patterns to users' acceptance of them. Relevant research papers are summarized in the following sections.

How Prevalent are Dark Patterns in User Interfaces?

Researchers have explored the pervasiveness of dark patterns and showcased that they are widespread across the internet as well as in mobile apps [26, 39, 42, 45, 52]. In an automated analysis of 11k worldwide shopping websites, Mathur et al. identified that 1,254 sites had issues, totaling 1,818 instances of dark patterns, with a higher likelihood of popular websites incorporating them [52]. Similarly, other researchers found that more popular websites were positively associated with four dark patterns and negatively associated with only one. In their automated detection of dark patterns in cookie consent requests, they found 3,744 different dark patterns in 2,417 consent dialogs [42]. Gunawan et al. conducted a study comparing the pervasiveness of dark patterns among different modalities. They found that apps have different dark patterns and tend to have more than their corresponding websites [39]. In an automatic walk-through of 240 popular free Android apps to detect malicious designs, Di Geronimo et al. found that 95% of the apps contained at least one dark pattern [26]. In another study focusing on devious ad content, MadDroid identified that approximately 6% of the tested 40,000 Android apps included such devious ad content [45]. Furthermore, dark patterns are not only

incorporated in screen-based interfaces but also in physical artifacts [18].
In a study by Mildner et al., they analyzed four of the most widely used global social network services, namely Facebook, Instagram, TikTok, and Twitter and found 44 different dark pattern types applied. It showed that Facebook employed the widest variety of dark patterns, applying a total of 41 different types. Following closely behind is Instagram with 39, TikTok with 37, and Twitter with 35 different dark pattern types [56].

Conti and Sobiesk already predicted an intensification in malicious interface techniques due to pressure from the marketplace and the fact that users get desensitized to these aggressive techniques in 2010 [21]. According to Di Geronimo et al., some designers copy the designs of others, which may explain the current widespread use of dark patterns [26]. Another reason for the pervasiveness of dark patterns is that companies simply know that they are effective in manipulating users [46].

Are Users Aware of Dark Patterns?

Dark patterns are so widely used that some users have gotten accustomed to them, thinking that they belong to the everyday interaction flow. Di Geronimo et al. use the term *dark pattern blindness* to describe the phenomenon of users not seeing dark patterns. In a study, they showed that users are generally dark pattern blind, with only one-quarter finding a malicious design in an app with dark patterns [26].

In contrast, other research shows that users are aware of the existence of such manipulative techniques [4, 10, 49, 50] while sometimes not being familiar with the term dark pattern [50]. Bongard-Blanchy et al. conducted a survey of whether users are able to recognize dark patterns. The findings indicate that 59% of the participants successfully recognized five or more dark patterns among the nine interfaces, with *Confirmshaming* and *High-Demand/Limited Messages* being easier to detect than others, like *Hidden Information*. The results illustrated that people are aware of design manipulating their behavior but were uncertain about the potential harm caused by it [10]. Research has shown that the level of education [10, 46] and the knowledge about dark patterns [26] influences the ability to detect such malicious design patterns.

Researchers request solutions to raise users' awareness about dark patterns [26] since study results suggest that raising users' awareness of the existence of dark patterns helps them identify such techniques [10, 54]. However, researchers also argue that raising users' awareness about online manipulative attempts does, on average, not impact the influence these designs have on users [10, 50]. While users are able to recognize dark patterns, they do not always understand their harm. Therefore, other countermeasures are needed [10].

How to Counteract Dark Patterns?

According to Schäfer et al., different dark patterns require different countermeasures [72]. Researchers tried to raise user awareness and created automatic detectors as well as other countermeasures.

Raising User Awareness: Researchers have looked into mechanics to help users recognize dark patterns. Aagaard et al. introduced a badge system, which is used for transparency among the utilized dark patterns within an app. Icons representing different kinds of dark patterns should be added to the download pages in the app stores to inform users that specific mechanisms are incorporated [4]. Bösch et al. introduced different countermeasures to seven identified privacy dark pattern types, ranging from including third parties that guide users through complicated settings to change their privacy preferences to raising user awareness [16]. Mathur et al. suggested using their automated detector of text-based dark patterns for further research [52], which Schäfer et al. did by introducing six different visual countermeasures to deal with three types of dark patterns detectable by the detector. They surveyed 40 people, prompting them to rank and rate the different countermeasures illustrated by prototypes on 7-point scales assessing usability, clarity, safety, efficiency, and helpfulness. The results showed that users want to have information about the presence of dark patterns but dislike visually cluttered interfaces due to the impact on their productivity [72]. In a literature review, Bhoot et al. identified five relevant variables for users to identify dark patterns: frequency of occurrence, trustworthiness, level of frustration, misleading behavior, and physical appearance of the interface [49].

Blaming Companies and Designers: Simultaneously, there are platforms that present and blame companies applying dark patterns, such as the official deceptive.design[1] website by Harry Brignull [14]. Additionally, people, including UX design practitioners, use social media platforms such as Reddit[2] [18] or X (formerly Twitter)[3] [31] to share their annoyance, discomfort, and frustration with manipulative design by posting examples. Posts on X related to dark patterns are tainted with negativity, either to raise awareness or to publicly blame companies [31]. On Reddit, the subreddit "asshole design" also includes manipulative designs in physical interactions, which tend to limit the functionality and usage of certain products, whereas dark patterns trick users into utilizing a particular service [18]. These posts and conversations show the concerns about and interest in discussing ethical issues related to dark patterns. Greenberg et al. assume that good companies would be able to turn a dark pattern into a good pattern which would balance company, client, and user benefits [38].
Contrary to this approach of counteracting dark patterns, Redström offers a different perspective and suggests that although the designer makes important decisions and determines the design, the user makes decisions as well. This said, the user can accept and also disregard certain things. There is an interaction between them, where the designer suggests something and the user accepts, rejects, or modifies it [66].

Automatic Dark Pattern Detectors: In order to develop automatic dark pattern detectors, researchers employed different approaches, such as using textual analysis [52], computer vision and natural language processing techniques [51], object recognition [45], or different

[1] `https://www.deceptive.design/types`, last accessed: 2024-08-07
[2] `https://www.reddit.com/`, last accessed: 2024-08-07
[3] `https://x.com/`, last accessed: 2024-08-07

approaches each tailored to the detection of one pattern [42].
Mathur et al. developed an automated technique to find dark pattern instances in the E-Commerce sector by exploring product pages, cart pages, and checkout pages. By considering visible HTML elements that contain text, their data set consisted of a list of segments with their HTML Element type, element text, dimensions and coordinates on the page, and style attributes involving text and background color. Then, they conducted an analysis focused on text, which limited their detection to text-based dark patterns [52].
Mansur et al. implemented AidUI, which automatically detects, classifies, and localizes ten unique dark patterns. In pursuit of this objective, they employ computer vision and natural language processing techniques that spot visual and textual characteristics of dark patterns in application screenshots. Specifically, AidUI inspects different user interface screenshots to identify (co-)appearing visual and textual cues that signify the existence of various dark patterns. It is limited to detecting 10 dark pattern instances with an overall average precision of 0.66%. The evaluation also showed that some dark patterns are more accurately detected, whereas others were not detected as often due to their varying implementation [51].
Liu et al. invented MadDroid, which automatically detects devious ad contents by using a combination of object recognition, character recognition, and deep learning. MadDroid, for instance, checks whether ad images contain the necessary x cross mark to close them. The different strategies to detect them are separated into different plugins, meaning one plugin for each devious pattern. Their efficiency relies on anti-virus scanners to flag malicious artifacts [45].
Kirkman et al. viewed ten different dark patterns found in cookie consent requests and analyzed the requirements for their automated detection. They carved out techniques in order to achieve it, including ones that make use of calculating brightness for detecting *False Hierarchy* patterns or others using Flesch-Kincaid for detection of the *Trick Wording* dark pattern. Subsequently, they implemented all these ten different techniques, achieving to detect dark patterns in consent dialogs with a 98.7% likelihood [42].
Curley et al. categorized dark patterns according to the possibility of automatically detecting them. They said some dark patterns are fully or partially automatically detectable, some are fully or partially manually detectable, and some are not detectable. The latter are not detectable as they are implemented in too many different variations. This category includes dark patterns like *Misdirection*, *Confirmshaming*, *Privacy Zuckering*, and *Bait & Switch* [25].

Other Countermeasures: Conti and Sobiesk mentioned potential countermeasures like pop-up blockers, text-only browsers and browser plug-ins [21]. Purohit et al. suggested that the very design principles responsible for addiction can also be used to alleviate those addictions. The hook model, commonly used in social media, represents a process of four steps to bond users: (1) a trigger leading to (2) an action that generates (3) a reward and creates (4) an investment. Subsequently, the researchers present nudges that work against these tactics by dismissing triggers, limiting user action, reducing rewards, and reducing investment. The *Default Nudge*, for example, turns off social media platforms by default. It interrupts the opening of a social media page by showing a screen where

users are told how long they have been absent, and they have to drag a slider to continue to the page. The dragging of the slider will become more challenging each time the user revisits the page. This leads to a mindful opening of this platform and prevents opening it out of routine. The authors *created friction* to facilitate a mindful interaction by the users. For example, users were asked if they wanted to see recommendations or not. And the *Pause-reminders* aimed to make users aware of their infinite scrolling through their feed. These new design ideas were tested by users through a browser extension and subsequently assessed on the System Usability Scale (SUS) and the AttrakDiff 2 questionnaire. The study showed that users believed they reduced their Facebook usage while simultaneously enjoying the time spent on Facebook more [65].

How Effective are Dark Patterns?

While companies can easily assess the efficiency of dark patterns by using their clients in A/B tests, it is more challenging for academics to assess it due to the lack of a database of study participants [46]. Additionally, it is very difficult to assess the effect on users on the level of one single element [58]. Still, there are a few papers investigating their efficiency.

Luguri and Strahilevitz assessed the efficiency of dark patterns in manipulating users and found that a mild dark pattern doubled the rate of participants accepting a company's preferred option, and an aggressive dark pattern almost quadrupled the rate of acceptance [46]. Similarly, Utz et al. tested if the number, design, or pre-selection of choices in a cookie consent request impacts users' decisions. Findings showed that the more choices, the more likely participants were to decline cookies. Results also showed that highlighting an option increased acceptance rates by 29.9% on mobile devices and by 27.5% on desktop devices. The pre-selection of choices led to acceptance by three in ten mobile users and one in ten desktop users, while almost no users ($\leqslant 0.1\%$) accepted the option when it required an opt-in choice [79].

Monge Roffarello and De Russis explored dark patterns that grab users' attention and get them hooked. One of their identified dark patterns was social investment, which refers to the display of social metrics, such as the number of comments, followers, or similar indicators. In a 3-week experiment using a browser extension to regulate social investment on a Facebook feed, they found that reducing this dark pattern led to a reduction in daily access and in the time spent on the page. Even though participants could tell that they used Facebook more actively and less intensely, only two out of seven participants recognized the changes. None of them experienced the browsing with the browser extension as problematic or negative [57]. In another study, participants were split up into a control group and a test group. The test group was interrupted every five minutes to ask if they wanted to continue scrolling, which can be considered as a *Nagging* pattern. While the control group spent an average of 19.7 minutes on Facebook, the test group averaged 9.2 minutes, demonstrating that this prompt led to a reduction of 50% of time spent. This, in a broader sense, also means that infinite scrolling manipulates users to spend more time on social media [60].

Benartzi et al. compared the effectiveness of nudging policies with standard policies by a literature review. Their data pool consisted of a variety of academic research that study interventions for influencing user behaviors to desired outcomes. Their focus was on how efficiently nudges achieve their goals as well as their cost-effectiveness. They state that the decision of whether to invest in nudging depends on the comparison of the impact of each dollar spent versus the cost of traditional interventions. They analyzed four different topics: retirement savings, college enrollments, energy consumption, and influenza vaccinations. They found out that nudges helped employees increase their retirement savings by preventing procrastination in signing up for a retirement plan, nudges helped high school seniors prepare for college by reducing procrastination and anxiety about errors, in turn leading to more college enrollments among high school graduates, nudges helped people reduce their electricity consumption, and nudges helped increase the number of flu shots. Nudges are usually more expensive than traditional tools, but there are also nudges that can be effective and cheap, generating high impact per dollar spent. For instance, their results illustrated that an average of 0.050 kWh of electricity was saved per dollar invested. The authors reflect that in some cases, nudges are not the best tool in achieving policy goals, thus, appealing for more investment in behaviorally informed policies, prompting organizations utilizing nudges to share data for further investigation, as well as prompting scientists to measure relative effectiveness of nudges. They conclude that governments should invest more in nudging since it can solve policy problems. [9].

Mathur et al. summarize that dark patterns can reduce three different kinds of individual welfare: financial loss, invasion of privacy, and cognitive burden. Moreover, from an individual autonomy perspective, some dark patterns maximize user engagement and enable addiction by a choice architecture influencing users to make choices they would not have done otherwise, denying a user choice, hiding available choices, or burdening the exercise of choice [53].

However, not all dark patterns are equally effective. Different types of dark patterns can affect users differently. While aggressive dark patterns can quadruple the acceptance of companies' desired options, they can also lead to customer backlash. However, mild dark patterns are more accepted by participants. Therefore, countering the subtle uses of dark patterns is more important [46]. Similarly, different groups of people react differently to malicious design interfaces. While these interfaces are usually not a problem for technically savvy people, they can become problematic and risky for others [21].

Why are Dark Patterns Effective?

Dark patterns make use of cognitive biases [52] and psychological insights [50] in order to manipulate users. One of these psychological insights provides that people do not read pages but only scan them [50]. For instance, users just click on anything they consider as most likely to eliminate the banner asking for cookie consent requests without understanding what they gave consent for [24]. This was proven by a study where only

10% read the cookie consent request, almost half of the participants skimmed it, and 40% did not read it at all [37].

Additionally, users appreciate defaults [50]. In a study, participants were asked by one of three different question designs whether they would be organ donors. One question design was an opt-in condition, where the default was not to be a donor, and they were asked to confirm or change their status. The second design was an opt-out condition, where the question was the same but the pre-selected default was to be a donor. The third question was a neutral condition where nothing was pre-selected. The results showed that the donation rate was twice as high when participants had to opt-out or with the neutral condition instead of opting-in. The authors also analyzed actual numbers of donations to assess whether the increased agreement rate also turned into an increased donation rate. The results showed that the opt-out option leads to a 16% increase in actual donations [40].

Moreover, people use the same option as they always do [37] and prefer to use easy and quick processes, which is seen by the selection of a button to close cookie banners. Only a few people will use a "manage cookies" link, where they do not know the steps ahead and how long it will take them to continue their intentional task, if there is also an "accept" button, which will immediately close the cookie banner [24].

While Luguri and Strahilevitz found in their study that dark patterns can lead to user disengagement [46], other research has shown that dark patterns did not have any effect on site revisitation. Thus, participants were still willing to visit websites, even though they knew there were dark patterns incorporated [17]. Similarly, in another study, participants were to some degree aware that companies misled them by incorporating dark strategies, but they accepted them as part of the internet experience. In their experiment, five out of ten participants provided their credit card details to Audible in a situation that should not necessitate such information. The authors suggest that the recognition of a reputable brand can lead people to use a website even when it requires sharing personal information in unnecessary contexts. Participants associated Audible with the well-known brand Amazon. Moreover, the authors argue that people use websites utilizing malicious techniques if they provide enough motive [49]. Maier and Harr elaborated on three different reasons why users continue to utilize particular services despite knowing and being concerned about their private data: (1) people are not aware of how their data is used, (2) people are willing to share data if the gained benefit exceeds the perceived risk, and (3) people feel the situation being out of their control as the persuasive monitoring is unavoidable. They foresee a future in which deceptive behavior is accepted and normalized [50]. Another reason could be the user's lack of motivation or opportunity to resist them, as Bösch et al. suggest [16].

People also use social media even though they know that their behavior is influenced by such platforms [50]. In a study by Bongard-Blanchy, most users mentioned Social Media services to be harmful by influencing the users [10]. However, research showed that users are aware of Facebook using tricks but still use it as there is no other platform where they can find information about other people more easily [49].

2.1.2 Types of Dark Patterns

The existing literature presents a variety of different dark pattern taxonomies, each crafted with distinct objectives [13]. Back in 2010, when the neologist Harry Brignull coined the term dark pattern, he only intended it to raise users' awareness and, thus, chose names that people would remember. Therefore, he did not elaborate on any strategies or a specific taxonomy [13]. Simultaneously, Conti and Sobiesk defined a taxonomy of eleven malicious interface techniques [21]. Since then, the literature on different dark pattern taxonomies has broadened. While certain categorizations center around particular domains like E-Commerce [52], Gaming [82], or Social Media [27, 56, 57], others classify dark patterns within specific contexts, such as in proxemic interactions [38]. Other taxonomies center their focus around dark patterns aiming a particular goal, such as intruding on users' privacy [16] or making users addicted to the system [57, 58, 60], and still, others create a general taxonomy, utilizing a broader approach [36]. There are even papers that highlight the diverse literature of dark pattern taxonomies and try to elaborate a new uniform [35]. Table 7.1 in the appendix provides a list of popular dark pattern taxonomies.

Approaches to Collect Data for Taxonomy Development

The researchers used different approaches to create the data pool for their taxonomies, including gathering practical examples found in practice, collecting examples from literature and practice, or basing their taxonomy on existing ones.

Creating taxonomies from practical examples: Several taxonomies have been developed from real-world examples found in practice [14, 21, 27, 38, 52, 56, 57, 82]. Since 2010, Harry Brignull gathered different examples of dark patterns found in different user interfaces, which accumulated to sixteen different types that are introduced and explained on the deceptive.design website[4] [14]. Mildner et al. recorded 16 hours of social media sessions to create a taxonomy of dark patterns employed on social media platforms. They analyzed four of the most widely used global social network services, namely Facebook, Instagram, TikTok, and Twitter, and identified 80 different dark patterns, 44 of which had already been described in the literature [56]. Similarly, Monge Roffarello and De Russis explored dark patterns that grab users' attention and get them hooked by self-monitoring their Facebook and YouTube usage for one week [57]. The guidelines by the European Data Protection Board describe dark patterns found in the entire life cycle of a user account [27]. Correspondingly, Mathur et al. created their taxonomy by automatically analyzing different shopping websites [52].

Creating taxonomies by combining practical examples and literature: Other taxonomies are combinations of examples found in literature as well as found in practice [16, 36, 60]. For instance, Gray et al.'s data set is a collection of examples indicating dark patterns in action shared by numerous journalists and practitioners. Over two months, they collected examples using search engines, social media sites, designers' blogs, commercial

[4] `https://www.deceptive.design/types`, last accessed: 2024-08-07

websites, and individual UX practitioners' professional websites, including the one by Harry Brignull [14]. This resulted in a corpus of 118 artifacts, which were then analyzed in a second step. In the first round of categorization, they used Harry Brignull's taxonomy as a set of codes and organized the majority of examples in this taxonomy. However, some examples could not be articulated in this taxonomy, which is why they iterated through the data set again using an open coding approach. The final taxonomy was built by a combination of these two iterations, using Brignull's taxonomy as a baseline but also focusing on the designer's intent [36].

Creating taxonomies from existing taxonomies: Some researchers developed their taxonomies based on other existing ones [35, 58]. Due to the wide variety of taxonomies in academic and regulatory literature without a uniform version, Gray et al. created a new taxonomy by integrating the ten most authoritative and commonly cited taxonomies from both these fields. At first, the authors listed all 245 patterns and identified high-level and low-level patterns. While a low-level pattern is referred to as a specific pattern, a high-level pattern is a category, strategy, goal, or intention. Then, they checked each original citation and combined the ones with similar definitions. Subsequently, patterns that shared common features were grouped into meso-level patterns. By completing this analysis process, they created a taxonomy of 6 high-level, 24 meso-level, and 34 low-level patterns. The six high-level patterns are *Nagging*, *Obstruction*, *Sneaking*, *Interface Interference*, *Forced Action*, and *Social Engineering* [35].

Relevant Taxonomies

In the following, the dark pattern taxonomy by Gray et al. [36] which influenced the creation of the newly developed taxonomy, as well as the three dark pattern taxonomies [27, 56, 57] that formed the basis of it are presented, chronologically ordered according to their publication date.

Gray et al.'s Taxonomy (2018) [36]: Gray et al. revisited Brignull's categorization [14] in 2018. They denoted that not all of these categories resulted from explicit and purposeful design intentions, but some of them were simply a result of poor design and thus should be called anti-patterns. Additionally, the authors indicated that the typology lacked a uniform way to describe the patterns, mixing context, strategy, and outcome together, thus, making comparison among patterns difficult. These two facts motivated them to elaborate on a new dark pattern taxonomy serving as a common vocabulary allowing for additional exploration. The structure of this taxonomy revolves around strategies and potential motivations of designers and contains five main dark pattern strategies, namely *Nagging*, *Obstruction*, *Sneaking*, *Interface Interference*, and *Forced Action*. With the exception of *Nagging*, all these categories of designer strategies include subcategories, some of which align with types identified by Harry Brignull [14].

Monge Roffarello and De Russis's Taxonomy (2022) [57]: Monge Roffarello and De Russis' taxonomy consists of five dark patterns namely *Recommendations*, *Autoplay*,

Pull-To-Refresh, *Infinite Scrolling*, and *Social Investment*. Its focus lies on encompassing dark patterns that grab users' attention and get them hooked.

European Data Protection Board (EDPB) (2023) [27]: This taxonomy encompasses 16 deceptive design patterns, in this thesis referred to as low-level dark patterns, grouped into six categories, in this thesis referred to as high-level dark patterns. The six categories include *Overloading*, *Skipping*, *Stirring*, *Obstruction*, *Fickle*, and *Left in the dark*. The taxonomy was developed as part of guidelines that serve as recommendations on how to evaluate and avoid these deceptive design patterns that violate General Data Protection Regulation (GDPR) requirements. Thus, the taxonomy focuses on dark patterns that violate users' data protection.

Mildner et al.'s Taxonomy (2023) [56]: Mildner et al's taxonomy includes five dark patterns in two design strategies, namely *Engaging Strategy* and *Governing Strategy*, which are found on social media platforms. The *Engaging Strategy* includes the *Interactive Hooks* dark pattern and the *Social Brokering* dark pattern, which both aim to keep users occupied and entertained on the social media platform as long as possible. The *Governing Strategy* aims to push users' decisions in a particular direction by encompassing the *Decision Uncertainty* dark pattern, *Labyrinthine Navigation* dark pattern, and the *Redirective Conditions* dark pattern. During their thematic analysis, the authors developed 22 codes grouped into five themes, representing the five discovered dark patterns. In this thesis, these codes are considered low-level dark patterns, while the five themes are considered high-level dark patterns.

2.1.3 User Perception of Dark Patterns

According to different studies, users encounter emotions such as sadness [7], disgust [7], anger [7, 46, 50], annoyance [49], upset [46], and frustration [49] when they recognize a dark pattern.

Gray et al. conducted a survey prompting participants to evaluate their felt emotions of past experiences while interacting with manipulative digital products. The participants had to indicate their agreement with a set of emotions on a 5-point Likert scale labeled from "Very Slightly or Not at all" to "Extremely." The list included the following negative emotions: distressed, upset, guilty, scared, hostile, irritable, ashamed, nervous, jittery, and afraid. The 169 participants most agreed with the emotions of upset, hostility, irritability, and distress, while they least agreed with feeling guilty and ashamed, indicating that they do not blame themselves for being manipulated. In follow-up interviews with nine survey respondents, one person stated that manipulative systems make them feel as if their creators prioritized the users' financial contribution over their humanity [34].

Maier and Harr used focus groups and interviews to explore how users perceive, experience, and respond to dark patterns. The nine participants, all of which were university students, were presented with a short overview of dark patterns along with examples and a definition during the focus group session. Then they were asked questions, such as what their first reaction to such patterns was. The researchers conducted follow-up interviews with 5 of

these participants later on. During these data-gathering sessions, participants mentioned *Toying with Emotion* as obviously recognizable and *Nagging* as annoying. They criticized dark patterns as sneaky, hidden, and intentionally implemented, and incorporating them would be a dishonest way of doing business. One participant stated that utilizing dark patterns would be a "shitty behavior" and lead them not to trust the service. These patterns made the participants feel angry, irritated, stupid, pressured, frustrated, worried, and stressed, and once they got used to them, they felt annoyed. However, the reaction depends on the harm of the pattern. *Nagging* and *Forced Action* were considered more acceptable by the participants since they are visible and not so dangerous besides being annoying. In contrast, patterns that are harder to detect were considered more dangerous. The authors state it is difficult to generalize the impacts and effects of dark patterns since participants' answers depend significantly on their previous personal experiences [50].

Bongard-Blanchy et al. researched user concerns towards dark patterns by prompting survey participants to rate six statements on a 5-point Likert scale and inviting them to elaborate on the answers in open-ended text fields. The results showed people were aware that dark patterns could influence their choices and behaviors, such as their spending behavior and content consumption. They believe these manipulative techniques cause psychological, physical, and financial harm. Others mentioned the cybersecurity/privacy harms and misleading information, which could influence people's opinions, in turn causing damage to society. In general, participants were not concerned about themselves being manipulated but about other, especially vulnerable, people getting manipulated [10].

In their dissertation, Calawen investigated the impact of dark patterns on the user experience in the travel context. They created two airline website prototypes, one with dark patterns and one without any manipulative technique, and tested them both by dividing the participants into two groups. In an experiment, 51 participants were asked to fulfill a task with one of the prototypes and subsequently were prompted to evaluate their experience on a UEQ, the SUS, and by answering open-ended questions. By comparing the evaluations of these two different prototypes, the results showed a significant difference on the *Attractiveness* scale. However, they could not observe any significant differences on the other scales. Thus, they concluded that dark patterns did not significantly affect the overall user experience [17].

Conti and Sobiesk did a survey on users' frustration when encountering malicious design and their tolerance towards these techniques. Participants found each technique frustrating, rating each technique at least a five on a 7-point Likert scale. Moreover, the authors did another survey and divided the participants into security-savvy experts and undergraduate college students. They found that undergraduate students tolerate great frustration caused by social networking sites, while experts do not. They believe this is due to the desirability of content on social networking sites to these users, whereas frustration with websites offering routine content such as news and weather is not as easily tolerated. The authors speak of a tolerance threshold [21].

Avolicino et al. gathered data on user emotions while encountering dark patterns. They filmed 50 participants while confronting them with seven different dark patterns and

subsequently analyzed the emotions obtained from facial emotion recognition. Among the recognized emotions were anger, fear, sadness, happiness, surprise, disgust, and no expression. The detected emotions were different according to the dark pattern [7].

2.2 Bright Patterns

After years of investigating dark patterns, there is a growing interest in exploring their opposite, referred to as bright patterns [37, 67, 78].

2.2.1 Definition

While literature does not yet provide a clear and established definition of bright patterns, a few researchers have taken some first steps. Hauke Sandhaus, a Ph.D. student in Information Science at Cornell University, describes bright patterns as *"persuasive user interface design functionality that prioritizes users' well-being and goals over their desires and business' objectives"* [67]. Other researchers use the term bright pattern to describe the opposite of dark patterns, dark patterns that are adjusted to the user's best option [37, 78].

Sandhaus argues that while bright patterns serve as the antonym of dark patterns, they extend beyond describing good design practices or the absence of dark patterns [67]. They are deliberately crafted to prioritize user long-term satisfaction over short-term gains [67]. The difference between long-term satisfaction and short-term gains can be illustrated by considering a person's long-term goals versus immediate desires. For instance, someone with long-term goals of being slim and physically fit might also have immediate desires to relax and be carefree. Experiencing these two conflicting feelings at the same time can create a dilemma [62]. In such a case, bright patterns should empower actions that support users' long-term goals.

The bright patterns incorporated into the interface are designed to align with users' goals rather than those of the company, even in situations where there is a conflict of interest between the parties involved. According to Sandhaus, there are three categories considered important for bright patterns, namely friction, transparency, and explainability [67]. This partly aligns with Graßl et al, who believe that freedom of choice and transparency are essential characteristics of bright patterns [37]. Freedom of choice can range from freely making one's own decision to limited decision-making up to no freedom at all. Transparency includes the categories of whether the information is clear, comprehensive, and/or complete [78].

In this thesis, the newly created bright patterns follow the definition of Graßl et al. [37] and Truong and Dalbard [78], whereby dark pattern techniques are inverted for the user's good. However, the whole topic is discussed in the broader sense, followed by the definition of Sandhaus [67].

2.2.2 Distinguishing Bright Patterns from Similar Terms

Recent literature presents a variety of similar terms that partly overlap with the term of bright patterns. Some of them will be described below, introduced by an image created by the author of this thesis to exemplify the definitions and enhance better understanding. As illustrated in Figure 2.1, the bright pattern highlights the opposite choice in a cookie consent request compared to the dark pattern.

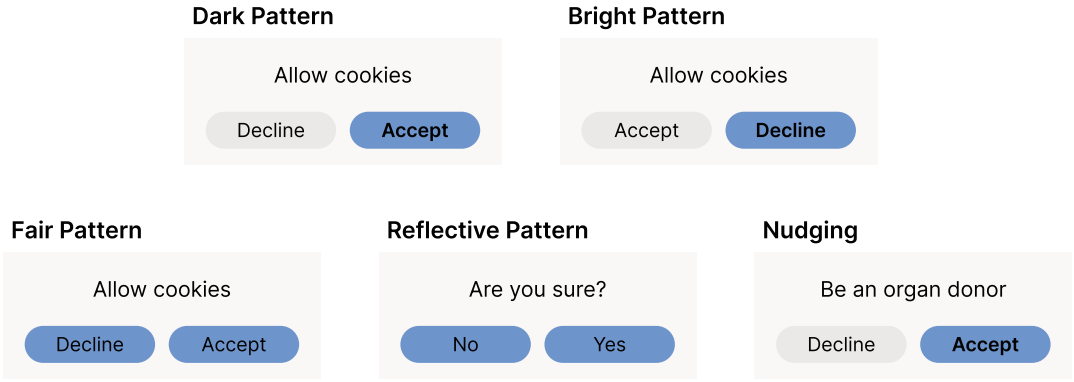


Figure 2.1: Examples of Similar Terms

Fair Patterns, sometimes also referred to as *Light patterns* [41], focus on users' free choice and autonomy. They foster System 2 decision-making, which is slower and requires more brain capabilities, but results in the correct solution since analytical strategies and cognitive efforts are applied. Fair patterns ensure this decision-making process by providing information that is easily accessible, balanced, consistent, and enabling [63]. As depicted in Figure 2.1, the fair pattern consists of two equally designed buttons when asking for cookie consent.

Similarly, *Reflective Patterns* aim at creating an environment in which users make reflective choices. Through friction, users are able to reflect on their actions, prioritizing introspection over efficiency, which is usually aimed for in user experience design. Users should reflect before, during, and after interacting with the digital technology, not only at decision time [76]. While fair and reflective patterns both support non-coercive design and enable decisions free from pressure and influence, they differ in the time they occur. While reflective patterns foster reflection at any time [76], fair patterns focus on free choices during decision-making processes [63]. Reflective patterns, as illustrated in Figure 2.1, create additional friction by asking for their confirmation of their decision to ensure users reflect on their decisions.

Nudges describe subtle modifications in the design of choices that influence people's behavior in a predictable manner without restricting options or significantly modifying incentives. Nudging involves shaping decisions in ways that guide individuals toward better outcomes while still maintaining their freedom of choice. For instance, placing

fruits at eye level is nudging, whereas forbidding junk food is not. Nudges are not mandates, and in comparison to bright patterns, a condition of them is to be easily avoidable [77]. Moreover, their focus is not on working against companies' interests, as bright patterns do[67]. Their focus lies on societal good, therefore the accept button is highlighted when asking users to become an organ donor, as depicted in Figure 2.1.

Persuasive Technology refers to any interactive computing system crafted with the aim of changing people's attitudes or behaviors. Fogg invented the term captology to name *"computers as persuasive technologies,"* which focuses on designing, researching, and analyzing such persuasive technologies [32]. Compared to the aforementioned terms, this is a broader concept.

2.2.3 Types

Since research on bright patterns is pretty new, literature lacks in a fully developed taxonomy. However, Sandhaus made first attempts by presenting and explaining fourteen different examples of bright patterns, namely *Slow down, Escape hatch, Simple Consent, Honest defaults, Nutrition labels, Persona profiling, Healthy alternative/5-a-day, Usage limits, Transparent recommender, Dumb it down, Data traces, Cost transparency, Outside my bubble,* and *Contrarian's companion* [67]. Moreover, on a website[5] people are invited to submit examples of bright patterns they have come across, which are subsequently presented in the Hall of Fame section.

2.2.4 User Research on Bright Patterns

Graßl et al. studied the influences of dark patterns vs bright patterns on the users' perceived control over their personal data and their decisions made in the realm of cookie consent requests. They ran the same experiment twice, first with incorporated dark patterns and afterward with bright patterns, while maintaining the same experimental conditions of both iterations for comparison. Participants were asked to look at eight news websites, each prompting them to give cookie consent when opened up. These consent requests incorporated the dark pattern concepts *Default Choice*, *Aesthetic Manipulation*, and *Obstruction*. In the second experiment, the authors reversed the direction of these design nudges to the benefit of the users. Instead of setting focus on the "Do Not Agree" button, which means no tracking, focus was set to the "Agree" button. Thus, in the second experiment of their study, they preselected the "Do Not Agree" button (*Default Choice*), color highlighted the "Do Not Agree" button (*Aesthetic Manipulation*), and replaced the "Agree" button with a "Manage options" button (*Obstruction*). During the study, users' decisions were tracked, and after reviewing all the websites, participants were asked about their felt control over their personal data and to what degree they deliberated on their decisions regarding each consent request separately.
Prior to the study, the authors had hypothesized that people would be more likely to choose the privacy-friendly option when the bright patterns are applied, which could be

[5] `https://brightpatterns.org/`, last accessed: 2024-08-07

verified for *Default Choice* and *Obstruction* but not be verified nor falsified for *Aesthetic Manipulation* as there was no notable effect. Overall, 93.8% of the participants agreed to the consent request in the dark pattern experiment, while 53.2% of the participants agreed to the consent when bright patterns were applied. However, they could not detect big differences between dark patterns and bright patterns regarding the perceived control over their personal data, as users felt little control in both experiments. Only *Obstruction* gave the participants at least a bit of control as they could change the settings [37].

Truong and Dalbard did similar research assessing users' decisions and users' perceived freedom of choice and transparency when opposed to dark patterns vs bright patterns in the E-Commerce sector. They also ran the experiment twice. One test run incorporated dark patterns and the other one included the contrasting bright patterns. The participants had to fulfill four different tasks on a graphical prototype built with Figma, whereby they were exposed to manipulative design patterns. The utilized dark pattern types were *Price Comparison Prevention*, *False Urgency Scarcity*, *Sneak into Basket*, *Trick Questions*, *Roach Motel*, and *Misdirection*. These were subverted to the users' best interest in the second experiment, focusing on higher freedom of choice and transparency. Thus, they showed prices per product (*Price Comparison Prevention*), did not show any false urgency scarcity messages (*False Urgency Scarcity*), opened a popup recommending another item instead of inconspicuously adding it to the shopping bag (*Sneak into Basket*), presented clear options (*Trick Questions*), embedded the link to delete the account in several locations (*Roach Motel*), and highlighted the "Delete" button instead of the "Never Mind" button (*Misdirection*).
The results of the study show that there is no difference in decisions made except for the process of account deletion, where *Roach Motel* and *Misdirection* were incorporated. However, bright or dark patterns change the way people perceive ethics. All of the participants preferred the prototype with the bright patterns instead of the one with the dark patterns. In addition, this prototype was perceived as providing higher freedom of choice and higher transparency than the other one [78].

By investigating design strategies to prioritize users' long-term goals over short-term desires, Ozkaramanli et al. found three options to motivate users to stick with their long-term goals and three options to demotivate them to not go with immediate desires. The former, intensifying long-term goals, includes adding new sources of pleasure to long-term goals, making the potential gains of long-term goals tangible, and creating enablers for long-term goals. On the other hand, weakening immediate desires, they recommend adding new sources of displeasure to temptations, making the potential losses of temptations tangible, and creating barriers to temptations. The findings were based on a four-week study investigating the motivations and feelings of 10 participants [62]. This aligns with another study investigating countermeasures against impulsive online shopping, in which users mentioned they had had success in the past when delays were incorporated [59].

While the aforementioned research offers valuable insights into aspects of users' perceptions, comparing dark patterns with bright patterns, the topic of bright patterns is

still under-explored, leaving a gap in understanding the overall user experience of such patterns as well as exploring these patterns in the context of social media.

2.3 Persuasion versus Manipulation

Even though bright patterns prioritize users' goals, they use dark pattern techniques, making them controversial due to the blurred line between persuasion and manipulation. This section explores the intersection of persuasive and manipulative design and presents relevant literature on how to define that boundary.

Intersection of Persuasive and Manipulative Design

Redström reviewed whether design is inherently persuasive. On the one hand, a design is normative, meaning it could take multiple other forms as well. Thus, the design brings people to see a particular point of view and shapes the way we think and live. But on the other hand, not every design is a matter of persuasion, some lack a designer's intention to persuade. Also, some products were not originally intended to be persuasive but were just designed and then turned out to be persuasive as they defined certain ways of usage [66].

Sánchez et al. outline that some dark pattern strategies align with some persuasion techniques, making these patterns hard to categorize. The authors characterize dark patterns as manipulative, hidden, and subverting vulnerabilities, while they characterize persuasion as transparent and allowing users' agency. The latter respects users' autonomy and avoids exploiting their vulnerabilities, while manipulation does not. In their study, designers linked manipulation to limiting user autonomy, pushing them into unwanted actions, or tricking and playing with them [75].

Distinguishable Examples

There are some examples that are rather clearly distinguishable. Mathur et al. state that there are design components that encourage users to make purchases in ways that primarily benefit shopping websites and retailers without fundamentally undermining user decision-making, which are not considered as dark patterns. For example, showing multiple images of one product or offering returns and exchanges for products [52]. Another example was provided by Schäfer et al., stating that a limited-time discount is not a dark pattern until it is an attention-grabbing fake countdown [72]. Truong and Dalbard exemplify it by some mobile applications that use advertisements in the app that can be turned off by using the fee-paying premium account. Although these advertisements distract users, they are not considered dark patterns, as they help the company generate revenue, and the user has the choice to spend money for a more convenient experience [78].

Difficult to Draw a Line

Sánchez et al. highlight the potential of traditional persuasive UX practices turning into manipulative designs [75]. While some dark patterns are blatantly manipulative, others are more subtle [78]. Based on their intentions and motivations, various types of dark patterns may not be equally "dark" but their common goal remains to manipulate and deceive users [36]. However, it is difficult to judge whether something is a negative experience or not [82], especially because some dark pattern instances, like *Forced Action* and *Nagging*, perform well in usability testing [36]. Additionally, ethical considerations are subjective, which means that a product can be considered good for someone, but bad for someone else [78]. In Maier and Harr's study, one participant stated that dark patterns would help them to decide faster, which can sometimes be good [50]. Additionally, it is difficult to determine the intention [82]. It may happen that an element is designed with good intentions but turns out to be dark later on, for example, if it is used for a different audience [36]. There does not exist a line that classifies patterns in good and bad [82], especially because there exists a gray zone [78]. Even designers have difficulties setting a concrete limit of manipulation despite being able to judge what is acceptable and what is not [75].

Non-Distinguishable Examples

Friction versus *Obstruction* leads to difficulties in deciding where to draw a line [75]. Cox et al. argue that small frictions, interaction points where users face difficulties during their interaction with technology, can have positive effects despite not aligning with traditional usability best practices that prioritize ease of interaction. They say that intentionally utilizing frictions can lead to mindful interactions, switching from a fast and automatic behavior (System1) to a slower, more deliberate one (System2). By adding one single step to a procedure that takes slightly longer than necessary can reduce the likelihood of errors as well as support health behavior change [23]. Contrarily, Gray et al. consider this as a negative *Obstruction* [36]. It is difficult to draw a line because they can be cumbersome for some users, but at the same time, for example, prevent users with problematic behavior from getting addicted [78]. Another example is the 30-day grace period in the account deletion process utilized by social media companies that facilitate to return back to the account if wanted. This can be perceived as good if the user regrets their deletion but can be perceived as bad if the user wants to delete their account immediately [56]. Sánchez et al. require friction to be analyzed in combination with the provided information in order to assess if it is manipulative [75].
Di Geronimo et al provided another unclear example, stating that some mobile apps offer the possibility of inviting friends to gain some in-app bonuses. Some users may perceive this as a legitimate business model, while others think it is inappropriate [26]. Another example that leads to a gray zone is what Mildner and Savino found on Facebook. They analyzed Facebook's interface design changes between 2004 and 2020 and found two changes that led to potential dark patterns. In particular, the log-out button and the link to privacy settings were relocated to a place hidden from the users. The authors argue

that it could be seen as an adaptation for aesthetics improvement, but it still results in obstacles for the users [55].

Attempts to Draw a Line

Zagal et al. state that the effect depends on the context in which the pattern is used, its implementation, intended audience and others [82]. Bösch et al. argue that in order to decide if a pattern is a dark pattern, it is important to take on a user's perspective, assessing how they feel and react to it [16]. Sánchez et al. made first attempts to investigate under which circumstances persuasive techniques become manipulative and harm users by studying designers' strategies to influence behavior. In order to persuade users to sign up for a newsletter, one participating designer suggested using friction and stickiness. The latter mechanism keeps repeating the call to action with the use of popups, sidebars, or top bars so that it is always present in the users' sight. They found that creating irresistible incentives in a hidden way by, for example, using emotional triggers can become manipulative. The authors exemplify this by stating that while giving convincing arguments to sign up for a newsletter is persuasive, it turns into manipulation if these arguments subvert vulnerabilities in a hidden way. Moreover, they argue that usability in itself can turn into manipulation if designers are aware of and make use of users' vulnerabilities. Furthermore, a different interpretation of the values of designs can lead to manipulation. For instance, if transparency and trust are used for only the purpose of attracting users, they might turn into manipulation. Their main findings are that whether something is manipulation or persuasion depends on the context in which the design takes place and the personal context of the designers. In other words, manipulative design can result from designers' intentions as well as from their circumstances [75].

Similarly, Susser and Grimaldi investigated automated influence by targeted advertising, digital nudges, and recommender systems. They argue that while dark patterns are not ethically correct, other forms of manipulation, such as digital nudges, are more complex. Even though they can encourage behavior change in a positive way, they can also harm users by invading their privacy, threatening their autonomy, and causing economic and epistemic harm. The authors invented questions for each of these topics that can be utilized in order to judge whether something is more good than bad. Some of these questions are: "Do people report feeling in control of their personal information?", "Do the targets of influence understand how the technologies targeting them work?", "What characteristics do influence technologies target? Do they seek to exploit people's weaknesses and vulnerabilities?", "Are people susceptible to conspiracy theories or other forms of disinformation in proportion to the amount of time they spend online, or in proportion to the relative share of news and other media they receive via targeted recommendations?". The authors request more systematic empirical investigations into the ethical and policy issues raised by automated influence technologies, focusing on their morally relevant aspects and their effects on individuals and society. The authors emphasize concerns regarding automated influence due to the current increase of artificial

intelligence and machine learning algorithms [74].

In the previously described study by Sánchez et al., the participating designers justified manipulation if the impact was positive. The designers who had to design for a Human Rights NGO did not consider some techniques manipulative if it was explained that they were for the organization and for the greater good, for example, projects to prevent war, hunger, or poverty. However, designers who had to design for a Fast Fashion company considered the same techniques as manipulative [75].

Also it depends on the area of application if a dark pattern is tolerable or not. Some patterns are designed to simplify the interaction, thus improving usability [57]. However, Gray et al. state that many persuasive strategies and even dark patterns can be used for good or ill [36]. Especially the attention-capture dark patterns can be perceived as good but if they impact the user's digital well-being, they can cause harm. It is crucial to determine the suitability of a particular mechanism in different contexts [57]. Participants tend to be more accepting of dark patterns when they are visible, when users have a choice, and/or when users find the interaction enjoyable and do not overlook anything significant [50].

Sánchez et al. introduce six design recommendations in order to not create manipulative design: supporting user goals, keeping incentives resistible, being careful with emotions, using friction for good, providing fair decision spaces, and being transparent [75].

2.4 Summary of Findings

- Dark Patterns are commonly incorporated into today's internet experience.
- Users are generally aware of dark patterns but uncertain about the harm they cause. In addition, users get, to some degree, accustomed to them.
- Researchers tried to raise user awareness, blamed companies and designers for creating dark patterns, and developed automatic detectors to counteract dark patterns.
- Dark patterns are effective in manipulating users, being able to quadruple the rate of participants accepting a company's preferred option.
- Dark patterns make use of cognitive biases and psychological insights.
- There are different taxonomies of dark patterns, each focusing on a different topic and developed by different data collection approaches.
- When recognizing a dark pattern, users encounter emotions such as sadness, disgust, anger, annoyance, upset, and frustration.
- While dark patterns aim to achieve business goals, bright patterns prioritize user goals over user desires and business goals.

- Bright patterns lead to higher freedom of choice and higher transparency but do not lead to higher perceived control over users' personal data. Moreover, study participants preferred bright patterns over dark patterns.
- Whether something is considered persuasion or manipulation depends on the designer's intentions and motivations, as well as the context in which it is applied.

CHAPTER 3

Methodology

3.1 Research Structure

This thesis aims to repurpose dark pattern tactics as bright patterns to benefit users and explore the topic of dark and bright patterns from end-users' perspectives. In pursuit of this, I employed both qualitative and quantitative research approaches, involving a literature review and a survey as well as the design and assessment of ideas to address the research questions. Figure 3.1 presents the methodological process of this thesis to answer the following research questions.

- RQ1: What are known problematic design patterns that manipulate users on social media, and how do users perceive them?
- RQ2: How could known dark pattern mechanics be repurposed as countermeasures against users' perceived social media problems?
- RQ3: How do users perceive the newly designed bright patterns compared to the corresponding dark patterns?

The whole process is divided up into two main parts. In the first part, I analyzed dark patterns in a descriptive and analytical research nature, analyzing the given facts solely and not influencing the parameters and variables in any way. While the descriptive research design aims to identify the characteristics, behaviors, or patterns within a specific group, the analytical research approach uses existing information to analyze a hypothesis that will subsequently either be accepted or rejected [8, 43]. As depicted in Figure 3.1, this part involved conducting a literature review on dark pattern taxonomies and a survey about dark patterns to answer the first research question. These two methods led to the development of a new taxonomy of deceptive dark patterns in social media, insights into how users perceive them, and a list of participants' perceived social media problems.

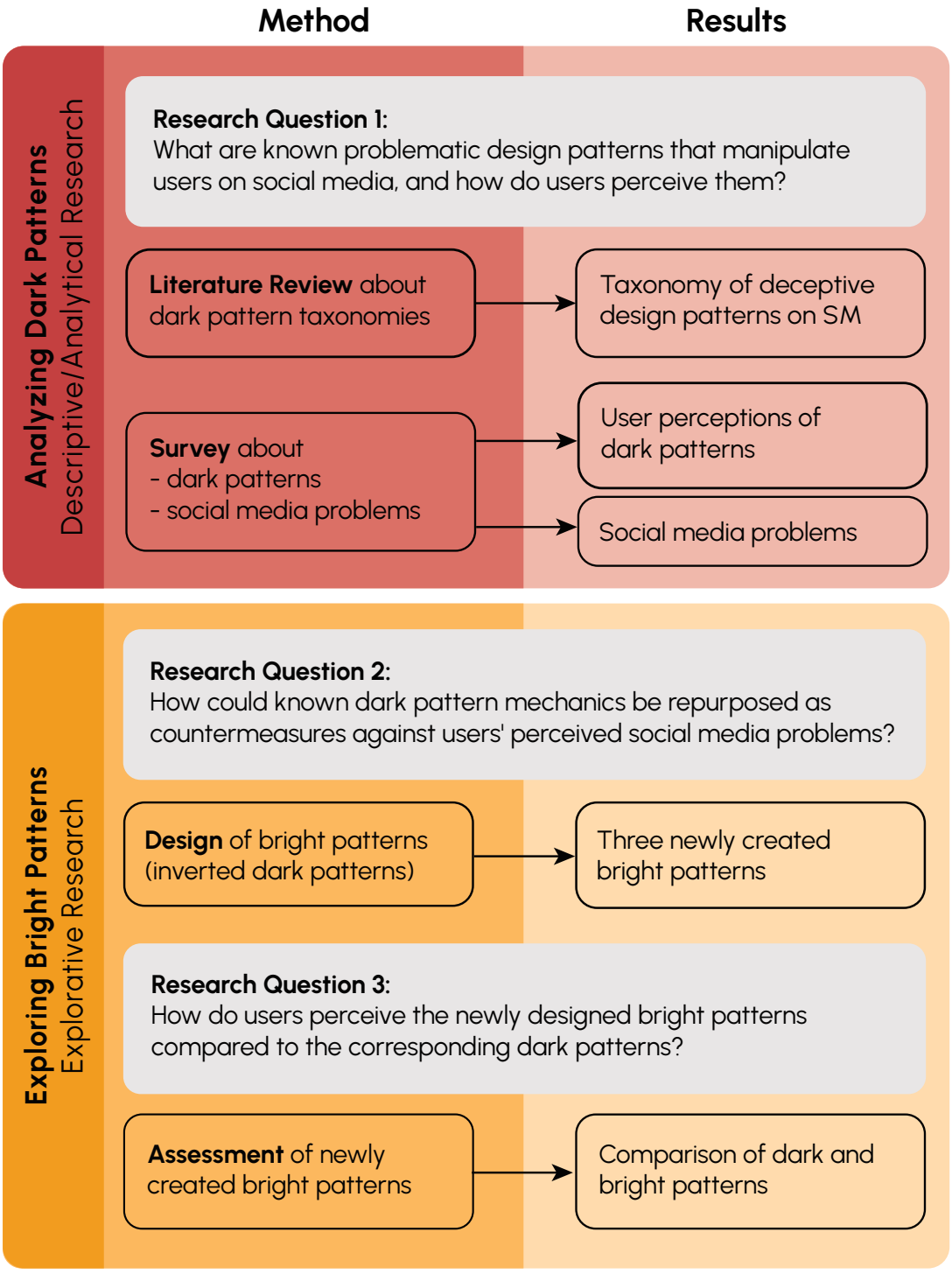


Figure 3.1: Methodological Overview: The analyzing dark patterns part involves a literature review and a survey to answer Research Question 1. The exploring bright patterns part involves designing bright patterns to address Research Question 2, followed by assessing these patterns to answer Research Question 3.

In the second part of the thesis, I took a more exploratory approach after collecting all the necessary information. The goal was to explore bright patterns and uncover a positive narrative of dark patterns. This explorative research design aims to reveal new insights, and therefore, it needs a flexible approach [8, 43]. This means that sampling design is non-probable, there is no pre-planned design for analysis, the collection of data is unstructured, and there is no fixed procedure [43]. Figure 3.1 illustrates that this part answered the second research question by designing bright patterns, resulting in the creation of three new bright patterns. It also answered the third research question with an assessment of these newly created bright patterns, leading to a comparison between dark and bright patterns.

The following sections explain the four methods in more detail: literature review (section 3.2), survey (section 3.3), design (section 3.4), and assessment (section 3.5).

3.2 Developing a New Taxonomy based on Literature

As Brignull states, each existing taxonomy has its own objective and is useful in its own way [13]. Therefore, it was important to create my own taxonomy of dark patterns found in social media. At first, an extensive literature review of the academic dark pattern taxonomies was conducted to find dark patterns that were already proven to be manipulative. In the second step, these papers were categorized as either relevant to this thesis or not. Subsequently, I developed a new taxonomy based on the three selected taxonomies.

3.2.1 Literature Collection

I started with a literature review, searching for dark pattern types and taxonomies in online academic libraries, both in general and also particularly in social media. At first I wanted to include more different papers into my data pool. However, I quickly realized that this is unnecessary and not sufficient since there are too many dark patterns that are not incorporated in the usual social media user journeys, for instance, the *Hidden Cost* dark patterns presented on the deceptive.design website[1]. Therefore, I had a detailed look into the taxonomies and then decided which of them to add to the data pool. In Table 7.1 in the appendix, I listed all these papers and one website and outlined the rationale behind excluding some of them. The selected taxonomies concentrate on dark patterns integrated into typical social media user experiences, not covering patterns in specialized features such as the Facebook Marketplace.

The final taxonomy of dark patterns created in this paper is a combination of three different taxonomies: two from academic papers [56, 57] and one from a regulatory document by the EDPB [27]. Table 7.2 in the appendix presents the used taxonomies with the different high-level and low-level dark patterns.

[1] `https://www.deceptive.design/types/hidden-costs`, last accessed: 2024-07-17

3.2.2 Literature Analysis

The identified taxonomies and dark patterns were analyzed with a process similar to affinity diagramming, with some aspects of thematic analysis. The process will be explained in detail in the following.

As a first step, I copied all 43 different low-level dark patterns (Table 7.2 in the appendix) along with their definitions into an empty document. Then, I started to read and reread the different definitions of dark pattern types. By this time, I had already read a lot of literature, especially after having already selected three papers for my data pool. This literature research helped me to familiarize myself with the body of literature, informing my analysis and ensuring a deep understanding of the data. Subsequently, I had a detailed look at dark pattern definitions that had same or similar names:

- Hidden in Plain Sight [27] vs Hidden In Plain Sight [56]
- Infinite Scrolling [56] vs Infinite scrolling [57]
- Pull To Refresh [56] vs Pull-to-refresh [57]
- Autoplay Content [56] vs Autoplay [57]

After checking whether they share the same definition, I merged them, resulting in 39 dark pattern types to analyze in the subsequent steps.

Affinity diagramming, sometimes also called affinity mapping, is a method to structure data and cluster them into groups by utilizing sticky notes and arranging them on a whiteboard [30]. Similarly, I created digital post-its, each representing one dark pattern by labeling it with its name, and randomly placed them on a whiteboard. While iterating through all the definitions, I organized the data and started to build groups of similar techniques. These groups of similar techniques are like themes in thematic analysis since a theme is a pattern of shared meaning [12], representing a significant aspect of the data concerning the research question [11].

In thematic analysis, there are two different ways to identify themes: with an inductive or a deductive approach. Inductive, also called bottom-up, means that the themes will be generated during analysis based on the data. Deductive, also called top-down or theoretical thematic analysis, means that the themes are chosen prior to the data analysis, and the data will accordingly be mapped to these themes. While the deductive approach tries to fit the data into a pre-existing coding frame, the inductive approach tries to find themes to the data [11].

I started with an inductive approach, freely grouping all the sticky notes. However, I quickly realized that my sorting was biased due to having read through so many taxonomies beforehand. I tended to categorize them according to different aspects I had read in papers. Therefore, I considered using a deductive approach. And since I did not aim to produce a taxonomy that is uniquely structured but rather one that frames the further process of my research, I found a deductive approach more appropriate.

I first attempted to map the dark patterns to the social media companies' goals: growing, generating data, and increasing user engagement [58, 68, 72]. However, I quickly realized that this is not sufficient since a dark pattern technique can be adapted to fulfill different goals. In addition, most of the dark pattern definitions did not include the goals or benefits of employing them. Therefore, I chose to rely on an already well-elaborated dark pattern taxonomy by Gray et al. as an *a priori* set of codes, similar to how these authors had used yet another taxonomy to create theirs [36]. On an online whiteboard, I organized the 39 low-level dark patterns labeled post-its within their taxonomy. Upon some iterations, I recognized that dark patterns that manipulate users into becoming addicted to social media did not fit into any existing theme. Thus, a new theme was needed. All six addictive dark patterns were grouped into one category, later called *Social Engineering*.
Other four dark pattern types were considered as themes (*Decision Uncertainty*, *Decision Governing*, *Forced Dialogue Interaction*, *Addictive Design*), and one dark pattern (*External Solution Search*) was excluded as it is defined by the action of users using search engines to find what they are looking for [56]. This was not considered a distinct dark pattern. The final taxonomy is presented in the next chapter in Table 4.1 and the affinity diagramming approach is depicted in the appendix in Figure 7.1.

3.2.3 Creating Mockups

In order for users to be able to evaluate the dark pattern types, I needed to present the patterns to them. Therefore, I decided to create mockups of the dark patterns in the next step. By creating my own mockups, I ensured that users would not see the companies that employ these dark patterns, which could potentially trigger a particular emotion and lead to biases. Thus, for each of the dark pattern types from the elaborated taxonomy, I created a picture or video mockup in a minimalistic social media app design with Figma[2]. In order to ensure a natural social media experience as well as to create appealing mockups, free stock photos were used from Pexels[3], a database of completely free stock photos licensed under the Pexels license. The mockups are re-creations of available examples presented in the literature. However, some dark patterns were mentioned in text descriptions solely without any visual presentation. In these cases, I based my design on the explanations.

3.3 Conducting a Survey about Dark Patterns and Problems

I conducted an online survey to investigate users' social media usage and their perception of certain features. In particular, I aimed to gather user opinions about three topics:

[2] https://www.figma.com/, last accessed: 2024-08-07
[3] https://www.pexels.com/, last accessed: 2024-03-03

1. My main goal was to find out how users perceive instances of dark patterns found in social media.
2. Additionally, I wanted to investigate what participants find problematic about social media and what changes they would suggest.
3. Moreover, I asked them about their social media usage and habits to potentially draw some conclusions.

Results from this survey helped me understand users' goals, laying the foundation for creating solutions and developing bright patterns. The approach of creating this survey will be explained in the following section.

3.3.1 Survey Creation

I started to gather survey questions which were inspired by and adapted from surveys discussed in other papers [19, 29, 34, 48, 50, 54, 55, 59, 61, 69, 81] and Websites [4,5,6,7,8,9]. Subsequently, all questions were categorized into the following blocks:

- Social media usage and habits
- Social media problems
- Social media dark patterns
- Ideas for changes
- Demographics
- Feedback and raffle

I used question types based on the aim of the question. Therefore, different types were used, each properly fulfilling the needs.

For the social media dark patterns section, participants were instructed to share their experiences of seeing such patterns. Other researchers investigated users perception of dark patterns with, for example, the UEQ [17], the SUS [17], emotion recognition [7], or mixed methods like focus groups and interviews to get more insights into participants feelings, thoughts, and experiences that they would not be willing to share in front of others [50] or a survey and follow up interviews to support the survey results with examples mentioned in the interviews [34], as described in the previous subsection 2.1.3. For this thesis, qualitative methods were not considered suitable due to their sensitivity to variations in user perception. Due to different levels of expertise, the same product can

[4] https://blog.surveyplanet.com/conducting-a-social-media-usage-survey-examining-trends-and-habits-through-survey-data, last accessed: 2024-01-14

[5] https://blocksurvey.io/templates/self-assessments/social-media-addiction-questionnaire, last accessed: 2024-01-14

[6] https://www.formpl.us/blog/social-media-survey-questionnaire, last accessed: 2024-01-14

[7] https://www.surveymonkey.com/r/6GX9GMT, last accessed: 2024-01-14

[8] https://candorhealthed.org/wp-content/uploads/2021/06/Candor-Health-Ed-Social-Media-Survey.pdf, last accessed: 2024-01-14

[9] https://www.surveymonkey.com/r/NZP7J8G, last accessed: 2024-01-14

be considered easy to learn by one person and complicated by another one. Consequently, there is a demand for quantitative user research methods, trying to gather feedback from as many users as possible. Additionally, Maier and Harr asked for quantitatively studying user perceptions of dark patterns as future research [50].

The SUS is one common method to quantitatively measure people's perception of a system. Participants have to evaluate their degree of agreement or disagreement with ten different statements on a 5 or 7-point Likert scale. This approach results in one single number between 0 and 100, representing the overall usability of the system [15]. The same assessment framework is used by the System Darkness Scale (SDS). However, the SDS assesses the darkness of a system and results in a final score that does not describe the usability but the users' perceived darkness of the tested system. In addition to representing this perceived darkness, the score facilitates comparisons of darkness levels across different systems. The SDS involves 5 questionnaire items, each worded negatively and focusing on the darkness of the system. For instance, one statement is "The system pushed me into spending more money than I originally anticipated". The result of this method does not represent how the user assesses each dark pattern but rather the overall system [80]. These two quantitative methods, SDS [80] and SUS [15], were dismissed from evaluating dark patterns in this work because they only result in one single output value, which was considered insufficient. Furthermore, the statements of the SDS are phrased negatively, which could result in response bias.

Consequently, the UEQ was selected for its holistic user experience evaluation, including both negative and positive adjectives. It is another common method to quantitatively assess how users subjectively perceive product features and the immediate impact these features have on them [44]. It consists of 26 items in six different scales to cover a comprehensive impression of the user experience towards a particular product [44]. These six scales are not independent. *Attractiveness* is the overall users' impression and a sum of the other five categories, which are mapped to hedonic and pragmatic qualities [71]. Hedonic and pragmatic qualities are both equally addressed in the UEQ [44] as research has shown that a product needs to fulfill both requirements to be considered attractive [71]. Whereas pragmatic categories, including *Efficiency*, *Perspicuity*, and *Dependability*, describe how users can accomplish their tasks, the hedonic categories, including *Stimulation* and *Novelty*, are considered as non-task oriented, questioning the originality and design of the product [44]. *Attractiveness* stands for the overall impression of the product, whether users like or dislike it. *Efficiency* characterizes if the user can achieve their goal without unnecessary effort. *Perspicuity* describes how easy it is to get familiar with the product. *Dependability* refers to the degree of the user's control over the product. *Stimulation* represents how exciting and motivating the utilization of the product is, and *Novelty* indicates to which degree the product is innovative and creative [71].

Each of the 26 items in the UEQ consists of a pair of contrasting adjectives separated by seven circles scaled from -3 (fully agree with the negative term) to +3 (fully agree with the positive term) [71]. In other words, the circles between the adjectives represent

gradations between the opposites. The user is asked to express their agreement with the adjectives by ticking one circle per item that best aligns with their impression.

Filling out the UEQ takes approximately 3-5 minutes, and because there are some scenarios where filling out the whole questionnaire is impractical, the inventors of the UEQ came up with a small version of it, the so-called User Experience Questionnaire Short (UEQ-S). It is a subset of the original UEQ, and consists of eight items, four items from the hedonic scales and four from the pragmatic scales [71].

Another advantage of the UEQ is the presentation of adjectives, which makes the evaluation easier for participants with difficulties describing their perceptions. In order to incorporate considerations for deception, the UEQ [10] was modified, which involved iterative steps described in the subsequent paragraphs. Following each step, I collaborated with Hauke Sandhaus, an expert in bright patterns, to review the adaptations and gather feedback. The resulting new questionnaire was called User Experience Ethics Questionnaire (UEEQ), and the steps to create it are explained in the following.

Steps to create the UEEQ

In the first step, I iterated through the 26 items of the UEQ and rated if they were applicable to evaluate dark patterns. In the next step, I iterated through the 15 dark patterns, looking at the whole UEQ again and making assumptions on which items the participants could have a strong opinion about. In other words, I looked at the mockup and guessed which items would be evaluated as a -3 or a +3. From this point forward, these assumptions are referred to as pre-survey assumptions. These two steps resulted in two different lists of items, which were then merged together. By doing so, I recognized that the *Novelty* scale is unnecessary for evaluating users' perception of dark patterns.

In another step, I compared the new questionnaire version to the UEQ-S to check whether it contained all of the UEQ-S' items. I reconsidered integrating the *Novelty* items of the UEQ-S which were *conventional/inventive* and *usual/leading edge*. However, as my aim was to keep it as short as possible to not overwhelm participants, I rejected this idea. Thus, the questionnaire consisted of 5 scales at this point.

After re-checking the adapted questionnaire, I noticed that I was missing some items asking for users' perceived manipulation and the techniques that dark patterns imply. A first idea was to add the item *manipulative/non-manipulative*, but I thought that manipulative might be too broad, resulting in users facing trouble evaluating whether they are manipulated or not. Therefore, I iterated through the 15 dark pattern mockups again and made assumptions about how each dark pattern could be described by one or more adjectives. The final list of pre-survey assumptions per dark pattern is presented in Table 3.1.

After this, I came up with four new items, namely *pressuring/suggesting*, *covert/revealed*, *non-addictive/addictive*, and *deceptive/benevolent*, each being their own scale. Thus, the

[10] `https://www.ueq-online.org/`, last accessed: 2024-08-07

Pattern	Assumptions
Nagging	• annoying • pressuring
Overcomplicated Process	• cluttered • confusing
Hindering Account Deletion	• deceptive • obstructive
Sneaking Bad Default	• annoying • covert
Expectation Result Mismatch	• complicated • obstructive • deceptive • confusing • unpredictable
False Hierarchy	• covert • obstructive
Trick Wording	• covert • confusing • deceptive • unpredictable

Pattern	Assumptions
Toying with Emotion	• unfriendly • pressuring
Forced Access	• pressuring
Gamification	• pressuring • addictive
Social Pressure	• addictive • pressuring • annoying
Social Connector	• pressuring
Content Customization	• enjoyable • addictive
Endlessness	• addictive • enjoyable • supportive
Pull To Refresh	• addictive • enjoyable

Table 3.1: Dark Pattern Pre-Survey Assumptions

final UEEQ consists of 9 different scales. It has two items in the scales *Attractiveness*, *Perspicuity*, *Efficiency*, *Dependability*, and *Stimulation*, and one item in the *Pressure*, *Addiction*, *Covert*, and *Deception* scales, resulting in a questionnaire of 14 items.

Finally, I translated the newly added items into German as the survey was offered in English and German to target a larger sample size. And as the UEQ aims at not forcing users to make abstract statements about their interaction experience [44], I added a *don't know/not applicable* option to each pair of adjectives. If participants found the feature unclear or believed that the adjectives did not describe it accurately, they could select this option. In addition, I randomized the order of the items and rearranged the pairs so that half of the items started with the positive term and the other half with the negative term. The final UEEQ is shown in Figure 3.2 in the layout of how participants saw it.

After I had planned the survey, I knew the requirements for the survey tool, which were handling different kinds of question formats, presenting pictures and videos of the size of up to 2MB, offering the user to choose between two different languages, and most importantly, not conflicting with the GDPR. Preferably, the tool would be free of charge

Figure 3.2: Newly Created User Experience Ethics Questionnaire (UEEQ)

as well. I chose the software tool SoSci Survey[11], which not only fulfilled the requirements but also offered the possibility to adapt the PHP, HTML, and CSS code.

3.3.2 Survey Design

The survey was targeted to everyone, whether they use social media, have used it in the past, or have never used it before. I refer to non-social media users as individuals who do not interact with any of the social media platforms presented in the survey on a daily basis. The length of the survey changed for non-social media users. These participants skipped the other questions and were redirected to questions on why they do not use these platforms and what modifications they would suggest to (re-)use social media. From this point forward, I will describe the flow of a social media user.

Before seeing the introduction screen, participants could choose between English and German as their preferred language. The introduction screen included a short description of the study's aim and framework, including the estimated duration. In addition,

[11] `https://www.soscisurvey.de/`, last accessed: 2024-08-07

information about data handling was given. Participants could only proceed with the survey after they had confirmed that they had read and understood these instructions.

In the following sections, each block of the survey is explained in the same order in which they were shown to the participants. The English version of the survey questionnaire can be found in the appendix in section B.

Social Media Usage and Habits

This part was dedicated to finding patterns of how users interact on social media and what their intentions behind using social media are.

The very first question consisted of a list of social media platforms, prompting the participant to indicate which social media platforms they engage with daily by selecting them. The list included the platforms: Facebook, Instagram, TikTok, Snapchat, LinkedIn, BeReal, Reddit, YouTube, Twitter/X, Pinterest, Twitch, and a *None*-option. If the latter was selected, the participant was treated as a non-social media user and thus redirected to questions on why they do not use these platforms and what modifications they would suggest to (re-)use social media.

Moreover, I asked them how much time they spend on social media on average each day and prompted them to select their main reasons for using social media from a predefined list. Then, there were six more Yes/No questions asking whether they had previously taken some actions on social media or not, like whether they had ever changed their privacy settings on any social media platform.

Social Media Problems

This section gathered users' perceived problems with two different kinds of question formats. Firstly, participants were prompted to evaluate different statements on a 7-point Likert scale. If participants were uncertain about a statement, they could choose the option *don't know*. These answers were excluded from the subsequent analysis. All the statements were grouped into four different problem groups: time and addiction-related, content-related, privacy and security-related, and well-being-related problems. Secondly, there was an open-ended question asking whether there is a particularly negative feature of social media they want to share.

Social Media Dark Patterns

In this part, users were tasked to share their experience with different dark patterns presented to them, each illustrated in a picture mockup, or video mockup if motion was involved. They were instructed to rate the dark patterns on the UEEQ, as previously explained in section 3.3.1. To prevent overloading and overwhelming participants with requiring them to evaluate 15 different dark pattern instances, only five instances were shown to each person. The selection of these five instances was randomized.

Ideas for Changes

In order to ask for users' goals, ideas for changes were gathered with an open-ended question format, asking *If you could talk to a social media platform's managers, designers, and developers, what modifications or new features would you suggest?*. This question was shifted to the back of the survey and was displayed after the dark pattern part.

Demographics

The last part asked participants for their demographics, such as gender, age, country of residency, employment status, and whether they consider themselves technically savvy or not. This part was shown to all participants, regardless of whether they were social media users or non-social media users.

Feedback and Raffle

All participants could leave any feedback, comments, or final remarks in an open-ended text field. Additionally, they could participate in a raffle to win an Amazon voucher.

3.3.3 Pre-Launch Setup

Before launching the survey, I conducted three test runs. The first was done with a technically savvy person working in industry, the second with a technically savvy person working in academia, and the last with a non-technically savvy person working in industry. Thus, I gathered important information about the clarity of the questions and could make adaptations after each test run, ensuring a clear and understandable questionnaire.

3.3.4 Recruiting Participants

After the test runs and adaptations, the survey was ready for launch. As mentioned in subsection 3.3.2, it was targeted to everyone, as I wanted to capture a wide range of impressions, and nearly everyone today uses social media. I started to recruit only a few people at a time to ensure I could make some changes if problems occurred. Luckily, there were no problems and so I recruited more and more people. Besides sending the survey to extended family and friends, I posted my survey on surveycircle[12] and in Facebook groups to target a broader audience.

3.3.5 Data Processing and Analysis

I went through the data corpus, encompassing all collected data, to compile the data set utilized for the analysis. I deleted all responses from test runs and from participants who did not indicate which social media platform they use daily and did not proceed to the last page of the survey. This raw data was then used for analysis. The only required data editing was translating the answers to the open-ended questions from German

[12] https://www.surveycircle.com/, last accessed: 2024-02-17

to English and not considering "don't know" answers to questions or subquestions. I analyzed the data using RStudio[13] and Excel[14]. The data analysis is explained in the following sections.

Social Media Usage and Habits

In a first step, I created bar plots of the data from these questions, which helped me find some patterns. Later, I used the data to find some potential correlations.

Social Media Problems

To analyze social media problems, I had two different data formats: data from Likert scale questions and data from open ended questions.

At first, I analyzed the Likert scale questions. I used the R package likert[15] to create 0-centered stacked barplots, which helped me to visualize the user responses and find outliers. Since all statements of the Likert questions were phrased differently, there was no one-size-fits-all threshold that could be used to verify whether a statement was a problem or not. The goal was not to find out how many participants perceive a specific issue as a problem, but rather to gather all different kinds of mentioned problems, meaning a qualitative analysis was preferred over a quantitative one.

Subsequently, my approach to analyzing the open-ended responses was similar to an inductive/deductive hybrid thematic analysis approach. This approach was valuable due to the different data formats. Similar to the example presented by Proudfoot [64], the inductive approach ensured codes from the data itself, which means that users' voices were valued as fully as possible. The deductive approach ensured a structured analysis of the data. This way, the open-ended responses could sometimes serve as intensification and clarifications to the statements from the Likert scale questions.

I started with an inductive thematic analysis approach to analyze the open responses. As already explained in subsection 3.2.2, inductive means that the codes are developed during content analysis [11]. This approach helped me to combine the data from the two different question formats. The phases of thematic analysis are as follows [11]:

1. Familiarizing yourself with your data
2. Generating initial codes
3. Searching for themes
4. Reviewing themes
5. Defining and naming themes
6. Producing the report

[13] https://www.rstudio.com/categories/rstudio-ide/, last accessed: 2024-08-07
[14] https://www.microsoft.com/de-at/microsoft-365/excel, last accessed: 2024-08-07
[15] https://r-graph-gallery.com/202-barplot-for-likert-type-items.html, last accessed: 2024-08-07

After reading through all the responses to familiarize myself with the data in step one, I then highlighted all relevant phrases in step two and labeled them with initial codes. Looking at all initial codes in step three, I grouped them and looked for themes. In step four, I changed from the inductive to a deductive approach. I looked at the statements from the Likert scale diagrams again and gave them a code as well. These codes served as themes for the thematic analysis of the open-ended responses. Then, I (re-)organized the codes to fit into these themes. I went back and forth between the whole data set, the Likert scale diagrams, the themes and the coded extracts, as analysis is not a linear process moving from one step to the next one [11].

Social Media Dark Patterns

Data gathered with the UEQ are typically analyzed by comparing the scale means [70]. In the same way, I analyzed the dark patterns based on their means. Therefore, I calculated overall means, considering all values from -3 to +3 across all character pairs, means per scale, and means per item. The former two serve as measures to compare dark patterns with each other, while the item means describe the dark patterns themselves.

Schrepp suggests a threshold of 0.8. Thus, if a value is $\geqslant 0.8$ it represents a positive evaluation, and if a value is $\leqslant$ -0.8 it represents a negative evaluation. The values in between these two thresholds represent a neutral evaluation [70]. These thresholds were not only used for scales, but also for single items. Thus, if, for example, a pattern scores $\geqslant 0.8$ on an item scale, it represents a positive evaluation, meaning the positive adjective describes the pattern. With the same thresholds, a pre-survey assumption is verified or falsified.

Furthermore, correlations of how participants evaluated the items of the different low-level dark patterns within a high-level dark pattern category were calculated. Therefore, I calculated the correlation coefficient between each pair of low-level dark patterns within a high-level dark pattern, considering each UEEQ item. Subsequently, I calculated the average correlation coefficient of all correlation coefficients within a high-level dark pattern by using the Fisher z-transformation[16]. This resulted in one number representing the degree of correlation among all low-level dark patterns within each high-level dark pattern.

Ideas for Changes

I aimed to map the ideas for changes to the social media problems detected in the previous step. Therefore, the responses to this open-ended question were analyzed with a similar process as the social media problems. I started with an inductive thematic analysis approach to code the open-ended responses. Afterwards, I searched for themes and created themes, resulting in feature and value themes. Subsequently, I aligned the

[16] https://matheguru.com/stochastik/korrelation-korrelationskoeffizient.html, last accessed: 2024-08-07

created themes with the themes from the social media problems. In other words, I checked which ideas from the feature themes would solve which social media problem.

3.4 Designing Bright Patterns

I explored dark patterns in order to create bright patterns that counteract social media problems, basing my ideas on findings from the previously conducted survey. Therefore, I started with an individual brainstorming session, writing down all the ideas that came to mind. I gathered post-its and pens and wrote down my goal of the brainstorming session, *"using dark pattern techniques to combat social media problems"*. Guided by the aim to generate as many ideas as possible, I did not restrict myself to feasibility or even reality [1, 73].

However, I quickly realized that using all fifteen dark pattern types to solve all social media problems would go beyond the scope of this thesis. Therefore, I decided to focus on a selection of three dark pattern types to combat three social media problems for the further part of the thesis.

Additionally, after I had familiarized myself with the dark pattern types in the first phase by mainly reviewing the literature, I felt the need for some more inspiration from other dark patterns in the wild. I wanted to look at more examples of how these techniques had been applied in practice, especially since the literature did not present as many examples as desired. Therefore, I browsed through some dark pattern databases[17,18,19,20] reviewing how the selected dark pattern techniques were applied. I dived deep into each of the three used low-level dark patterns and created mindmaps, grouping together similarities in design choices incorporated by the different instances. The entire mindmaps were too big to add to this thesis, therefore I generated smaller versions of them, each technique depicted by only one example. These can be viewed in the appendix in section C.

After this brainstorming, I felt the need for a more structured approach, leading me to the mash-up ideation method [2, 3, 73], which serves as a thought exercise, speeding up ones thinking and promoting creating bold ideas [3]. The idea is to combine different topics to create new ideas using the following steps. At first, the design challenge is to be formulated in a *"How might we"* statement. Subsequently, two broad categories that are unrelated to each other are chosen. At least one of the categories needs to relate to the design goal. Then, as many elements as possible are written down for each category separately. In the last step, the elements are mashed up, which means that items from each category are combined to generate ideas [2, 73]. In my thesis, the statement was *"How might we use dark pattern techniques to milder social media problems?"*, and the two categories were dark pattern techniques and social media problems requiring combating mechanisms. I broke down the big goal of finding a positive narrative of

[17] https://darkpatterns.uxp2.com/patterns/, last accessed: 2024-08-07
[18] https://www.deceptive.design/hall-of-shame, last accessed: 2024-08-07
[19] https://darkpatternstipline.org/sightings/, last accessed: 2024-08-07
[20] https://confirmshaming.tumblr.com/, last accessed: 2024-08-07

dark patterns to create healthier social media usage into smaller goals. Then I created a matrix, putting the perceived problems of social media listed as the columns and the selected three dark pattern types listed in the rows. Moreover, I divided my ideas into patterns that are integrated into the feed, as well as those integrated into the story or post-creation process. The final matrix of ideas is depicted in the appendix in Figure 7.5. Afterwards, I selected three ideas and created mockups. The mockups were designed with Figma[21] again, reusing stock photos from Pexels[22] and the design I had already created in the first part.

3.5 Assessing Bright Patterns

In order to assess the bright pattern ideas, I conducted a mixed-methods user study. The aim of the study was to gather information about participants' experiences and emotions when being opposed to these design patterns, and their acceptance towards them. In particular, the latter assessed whether participants believe the patterns are manipulative, and if yes, whether they think manipulation is okay. Additionally, I was interested in whether they could recognize that the different patterns had different intentions and whether it was clear that some had good and some had bad intentions. The upcoming sections present the study design, pre-study setup, and how I analyzed the data.

3.5.1 Study Design

The study was divided into two main parts: (1) participants evaluated the patterns individually in a survey setting, and (2) participants answered follow-up questions in an interview setting. This mixed-method study was chosen for several reasons. Firstly, the individual setting ensured that participants were able to evaluate the patterns by themselves without any biases due to the interviewer. Secondly, besides the reasons for using the UEEQ that were already mentioned in section 3.3.1, this questionnaire is of high value when comparing two products [70]. Thus, it facilitated the comparison of the bright patterns and the dark patterns. Thirdly, the interview was chosen to gather detailed information from participants that could not be captured with the survey [22], providing a deeper understanding of the domain of bright patterns. It was conducted immediately after the survey so that participants could still elaborate on their first impressions.

In order to create a comparison between the newly created bright patterns and their corresponding dark patterns, the study included not only the bright patterns but also the dark patterns that were mocked in the first part of this thesis (subsection 3.2.3). Thus, participants assessed three dark patterns and three bright patterns, summing up to six different patterns.

In the introduction to the user study, I explained the procedure and the consent form (Figure 7.6 in the appendix) and asked them to sign it. Additionally, I tried to create a

[21] https://www.figma.com/, last accessed: 2024-08-07
[22] https://www.pexels.com/, last accessed: 2024-03-03

welcoming and casual environment where participants could share their opinions. This was even more important due to the fact that the topic was about manipulation and ethics, which is prone to the prestige response bias or social desirability. The latter means that participants do not tell the truth but the response they believe is socially desired, while the prestige response bias describes the phenomenon of participants providing answers that increase their image [22]. In order to prevent this, I made clear from the beginning that all information provided will be treated with the utmost confidentiality, and that their responses will be used solely for research purposes. They will be analyzed anonymously and will not be shared with third parties. Additionally, I encouraged them to answer honestly by stating that I was aiming to gather honest responses and that there is no wrong or right answer; it is their personal opinion that counts. After this introduction, the subsequent part of the study was divided up into two parts, as explained below.

Survey

In the survey, users' perceptions of the six patterns were assessed individually. So to say, one pattern was presented and beneath it, users were prompted to evaluate it on the UEEQ (explained in section 3.3.1) as well as to answer the question *"How does it make you feel? Can you elaborate on the kind of emotions you are feeling when being exposed to this design element/feature."* Beneath was a large input field, not restricting their thoughts and opinions.

After that, participants were asked for their gender, age, home country, employment status, and whether they considered themselves technically savvy or not.

Interview

The interview was of a semi-structured nature since this type of interview is considered a good method to add depth to mixed methods research [5]. Thus, I had created an agenda for the interview and planned a set of questions but deviated from time to time [5, 22]. As suggested by Adams, I prioritized the most important questions and highlighted them on the interview guide [5] by using a red font, as seen in the appendix in Figure 7.7. The interview guide was based on questions other researchers had asked participants in similar studies before [20, 28, 34, 50, 61]. I came up with an interview guide where the questions were divided into four parts: General, Attractiveness, Perceived Efficiency, and Acceptance.

In the first part of the interview, each pattern was elaborated on individually. I displayed one pattern at a time and always started with the question, *"What are your thoughts on this design element/feature?"*. If necessary, I asked open-ended follow-up questions in order to let them elaborate on their answers.

After iterating and discussing each pattern individually, I asked some more general questions. I displayed an overview of all the mockups they had seen, asked them whether there was one feature that stood out, and if they had a strong positive or negative opinion

about any of them. After that, I also wanted to know their opinion about manipulation. I pointed at particular patterns they had categorized as manipulative in the first part of the interview and asked them whether they excuse such manipulative behavior under certain circumstances and if they can think of examples when such manipulative behavior is okay or even desired.

3.5.2 Pre-Study Setup

Before starting to gather data, I ran a test run with a social media user, which was considered the final step in refining the interview guide [5]. The focus was set on gathering information about the clarity of the questions, whether they are not biased, and getting a feeling of how long the study will take. After running the pilot study, I knew that the general question could lead the conversation into a variety of different directions and potentially already answer some of the other questions. Additionally, the participant gave the feedback that it was sometimes not clear what exactly needed to be evaluated. Therefore, I modified the text above each mockup to a scenario, starting with "Imagine...". In addition, I removed the question about the participant's feelings from the interview to the survey. This way, the participants were prompted with this question right after being opposed to them, ensuring they would catch the first impressions.

3.5.3 Recruiting Participants

As with the survey in the first part of this thesis, I did not target a specific group of people. The only requirement for participants was that they had not taken part in the survey conducted in the first part of this thesis, as the goal was to capture their first impression. I recruited students at Indiana University of Bloomington that I got to know at events organized by the international office. Therefore, most participants were international students.

3.5.4 Data Processing and Analysis

Unlike in the first part of this thesis, where a threshold of 0.8 was chosen to differentiate between neutral, positive, and negative evaluations, a threshold of 1.5 was selected to analyze the data from the UEEQ due to the low number of participants.

The interviews were voice-recorded and with Otter.ai[23] automatically transcribed. Subsequently, the interview transcripts were copied into a new text file and then inductively coded with thematic analysis. I followed the six phases of thematic analysis [11], as listed before in section 3.3.5.

Each interview was listened to twice. The first round was mainly to check the transcription and correct mistakes made by the software tool, and the second one was to focus on the spoken conversation and take notes for possible codes. Both these two iterations are considered the first step of thematic analysis, helping to get familiar with the data. Then,

[23] https://otter.ai/, last accessed: 2024-06-25

I highlighted all the statements relevant to answer my research question and added first versions of the codes to them. After this iteration of adding the first codes, I looked at all my codes, grouping them into themes. Once again, I went back and forth between the data set, the themes, and the coded snippets to code all the data.

As the interview consisted of one part of assessing the individual patterns and one part of a general discussion about manipulation, the codes for these different parts were treated differently.
The codes that describe the individual patterns were added to a table, with the dark pattern on the left side and the bright pattern on the right side. This way, the newly created bright patterns could be compared with their corresponding dark patterns. The table of codes from the dark and bright patterns can be found in the appendix in subsection D.3.
The codes for the more general part were simply listed under each theme. The final themes and sub-themes are presented in the following Table 3.2.

Themes	**Sub-Themes**	**Relevance**
Improvements per Pattern	-	Discussion
Attractiveness and Acceptance per Pattern	General like/dislike	RQ3
	Positive adjectives	
	Negative adjectives	
Manipulation	Definition	Discussion
	Acceptance	
Dilemmas	Social Media Context	Discussion
	User behavior, Contextual Knowledge	
	Others	

Table 3.2: Thematic Analysis Codes for Pattern Assessment

CHAPTER 4

Findings from Analyzing Dark Patterns

This chapter presents findings derived from descriptive and analytical dark pattern research. It includes a literature review of different taxonomies resulting in a new taxonomy, as well as a survey gathering users' perceptions of dark patterns.

4.1 New Taxonomy and Mockups

This section presents a new taxonomy of dark patterns, compiled after literature review. The final taxonomy organizes the various dark patterns common in social media, consisting of 15 low-level dark patterns mapped to 6 different high-level dark patterns, as presented in Table 4.1. While a low-level pattern refers to a specific pattern, a high-level pattern presents a broader strategy. The 6 high-level patterns, are *Nagging*, *Obstruction*, *Sneaking*, *Interface Interference*, *Forced Action*, and *Social Engineering*. The following sections provide an explanation of each high-level dark pattern, along with their corresponding low-level dark patterns, which are illustrated using the created mockups. Twelve out of the 15 low-level dark patterns were mocked in an image. The remaining three (*Nagging*, *Endlessness*, and *Pull To Refresh*) were mocked in a video since an image would not have appropriately presented the pattern.

4.1.1 Nagging

Dark patterns categorized as *Nagging* are small redirections or other annoying distractions that repeatedly show up and lead the users to do something other than their intentions. They consume users' time and patience and won't disappear until users give in and accept the provider's offer, suggestion, or request. Examples of this dark pattern strategy and type are pop-ups or audio notices [36]. This high-level dark pattern includes only one

low-level dark pattern, named identically. Figure 4.1 presents a screenshot of the video mockup, where a pop-up repeatedly prompts the user to turn on their notifications.

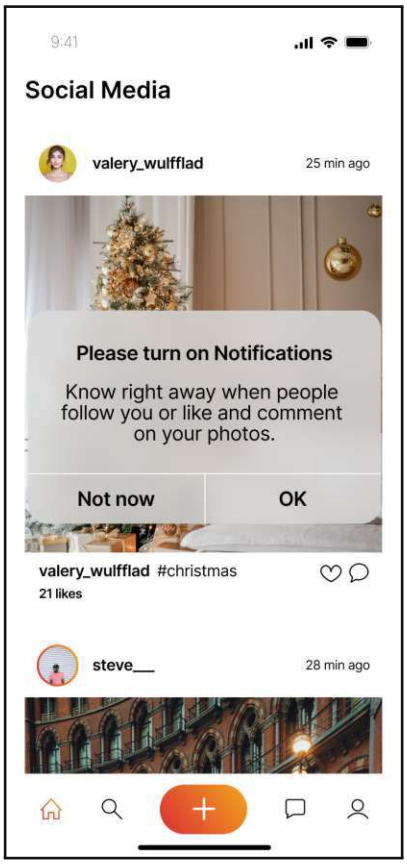


Figure 4.1: Nagging Dark Pattern Mockup

4.1.2 Obstruction

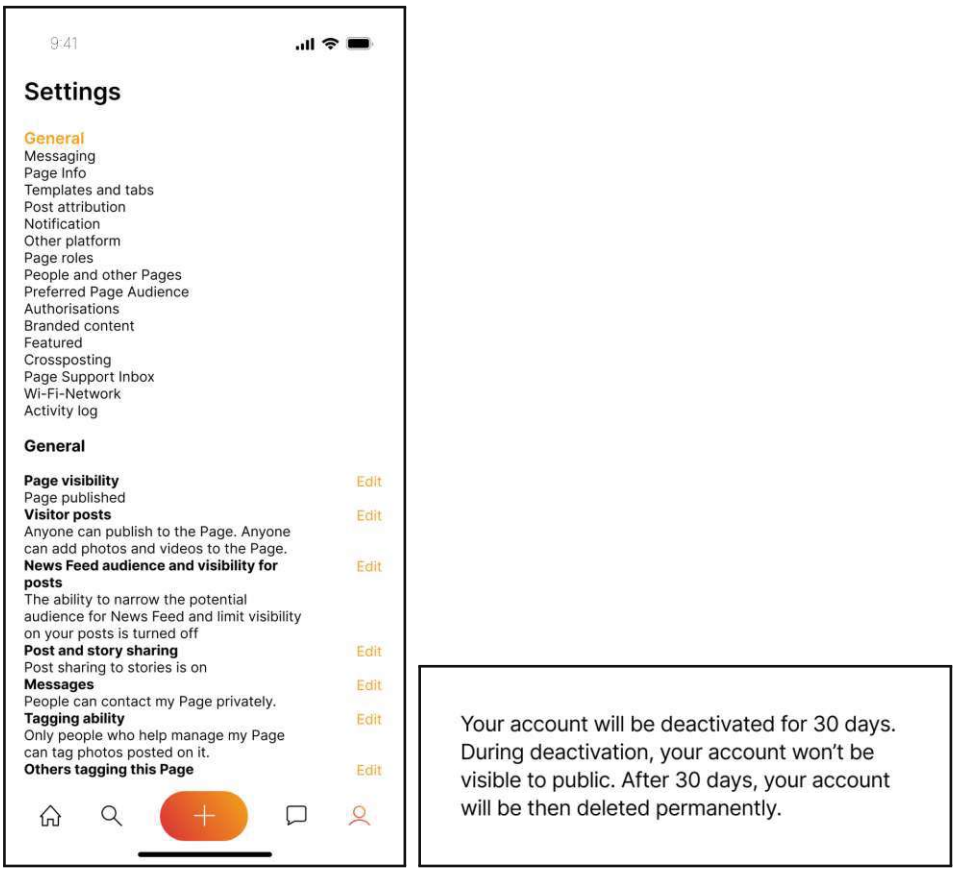
This strategy refers to dark patterns that interrupt users, impeding them from fulfilling the business' undesired tasks or taking a certain action. The objective of these patterns is to wear down users and encourage them to give up if the users' actions do not align with the business's revenue or growth goals. This pattern includes the following two low-level patterns [36].

Overcomplicated Process

This dark pattern makes situations intentionally complicated, incorporating unnecessary steps, too many options (in inconsistent ways), or a lack of clear hierarchy [27, 56]. The mockup, depicted in Figure 4.2a, presents an overwhelming, unstructured and cluttered overview of account settings.

Hindering Account Deletion

This dark pattern, also referred to as *Roach Motel*, describes situations that users easily get into but face trouble getting out of. It is applied if signing up for a subscription is possible within a few seconds, while the cancellation is cumbersome, sometimes including calling a customer service line [36]. Figure 4.2b displays information presented to the user after they had chosen to delete their account. It states that their account will be deactivated for 30 days before being permanently deleted.



(a) Overcomplicated Process (b) Hindering Account Deletion

Figure 4.2: Obstruction Dark Pattern Mockups

4.1.3 Sneaking

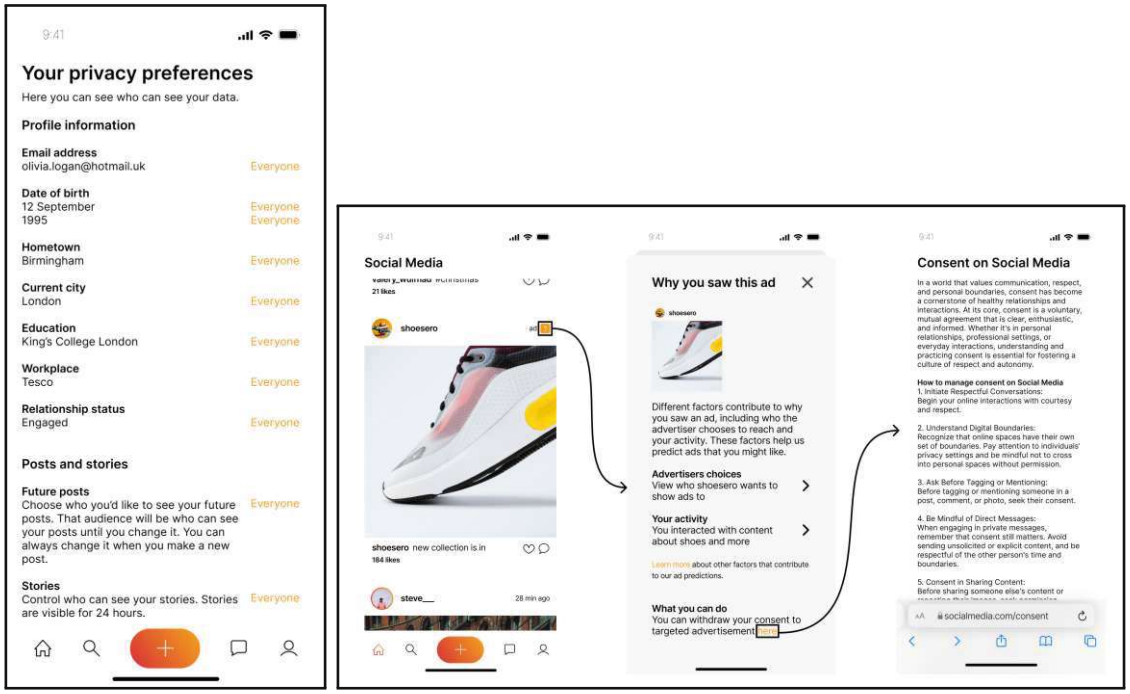
The strategy of *Sneaking* involves hiding, disguising, or delaying the display of relevant information to the user. The absence of this information leads to actions that the users would not have taken if they had the information. Relevant information is, for example, the undesired effect from a particular action. The following two low-level dark patterns are grouped into this strategy [36].

Sneaking Bad Default

This dark pattern includes a decision being made for the user, so they do not need to choose themselves. However, the decision is made in the provider's best interest [36] and users are sometimes not even aware of it [27]. Figure 4.3a presents a snippet of privacy preferences settings where each visibility option is per default set to Everyone.

Expectation Result Mismatch

This dark pattern, also called *Bait & Switch*, utilizes actions that seem apparent to the users but, in fact, trigger a different undesired action. For instance, a mobile game exploits muscle memory by strategically placing a button to purchase more moves in the same location as the button to start a new game [36]. The mockup presented in Figure 4.3b displays where the user is redirected upon clicking on two links. The user expects an option to withdraw their consent to targeted advertising but is instead presented with a page full of general information about consent on social media, without any option to withdraw.



(a) Sneaking Bad Default (b) Expectation Result Mismatch

Figure 4.3: Sneaking Dark Pattern Mockups

4.1.4 Interface Interference

Dark patterns utilizing this strategy manipulate the user interface in a way that specific actions are favored over others. This leads to confusion and limited discoverability of

possible options by the users [36]. This pattern includes the following three low-level dark patterns.

False Hierarchy

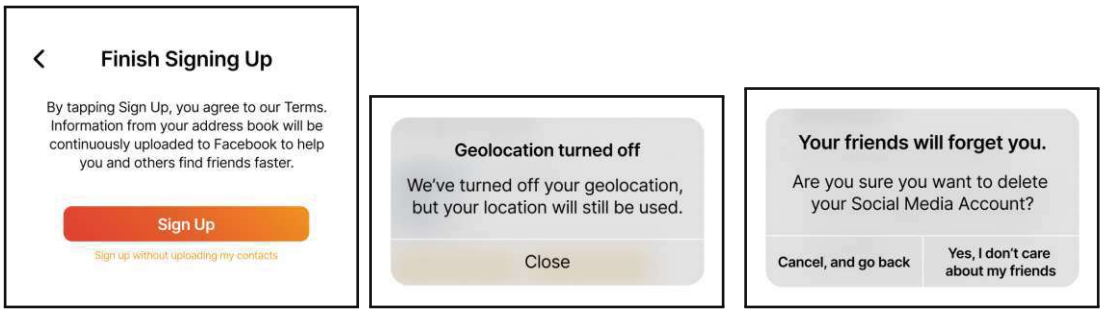
This dark pattern presents options that are equal and should therefore be equally designed but are differently designed so that the user thinks one option is the best or even the only possible option. For example, one option may be designed in gray, to appear disabled, but is actually clickable [36]. Figure 4.4a presents two options: One big orange button and another one in small orange font below.

Trick Wording

These are questions or statements that appear to mean one thing but, in fact, mean something else. This pattern takes advantage of perplexing language, like tricky wording or double negatives, to confuse the user. For example, a checkbox to opt-out rather than opt-in [36]. Figure 4.4b presents an information pop-up stating that the geolocation is turned off but simultaneously says that the location will be used.

Toying with Emotion

This dark pattern utilizes language, style, color, and other elements to elicit an emotion in the user, subsequently influencing the user to take a specific action. These elements might be cute or scary images, or a button text to decline an option could, for example, be, "No thanks, I'll have a microwave dinner tonight." [36]. The mockup depicted in Figure 4.4c presents a pop-up stating that users' friends will forget them if they delete their account and also that deleting their account would mean that they do not care about their friends.



(a) False Hierarchy (b) Trick Wording (c) Toying with Emotion

Figure 4.4: Interface Interference Dark Pattern Mockups

4.1.5 Forced Action

Dark patterns utilizing this strategy force users to do something in order to fulfill a specific task. Thus, users are only able to succeed if they do something else in return.

For example, in order to unsubscribe from a mailing list, a user is required to agree to allow the company to sell their information to other countries [36]. This pattern involves one low-level dark pattern, described in the following.

Forced Access

This dark pattern requires users to give access to special device hardware or local data in order to use some features [56]. Figure 4.5 depicts the process of giving access to one's photos in order to save a photo.

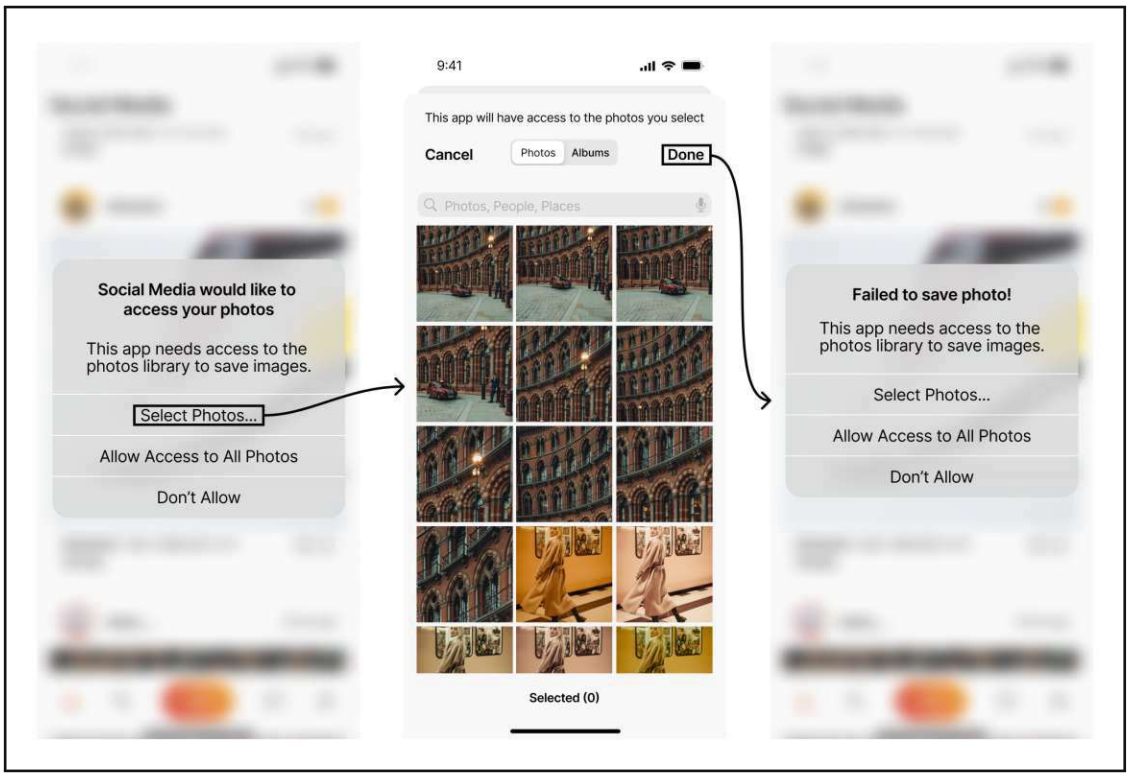


Figure 4.5: Forced Access Dark Pattern Mockup

4.1.6 Social Engineering

This strategy is all about bonding users to the platform as long as possible. These low-level dark patterns try to get users hooked on social media. There are six low-level dark patterns, as explained in the following.

Gamification

These are situations where specific aspects of a service can only be obtained through repetitive (and possibly unwanted) use of features within the service. For example, video

games grant points and level-ups if users repetitively play [36]. Figure 4.6a presents that the user is on a 31-day strike with another user and that this strike will reset if they do not continue the chat the day after.

Social Pressure

This dark pattern manipulates users by creating a fear of missing out [56]. Figure 4.6b depicts a notification saying that another person added a story and they should open the app in order to not miss out.

Social Connector

This dark pattern prompts users to give additional information to connect to friends [56]. It can also persuade users to invite friends and family to the system in order to use the service or by rewarding them with special benefits [36]. Figure 4.6c illustrates suggested accounts to follow and connect.

Content Customization

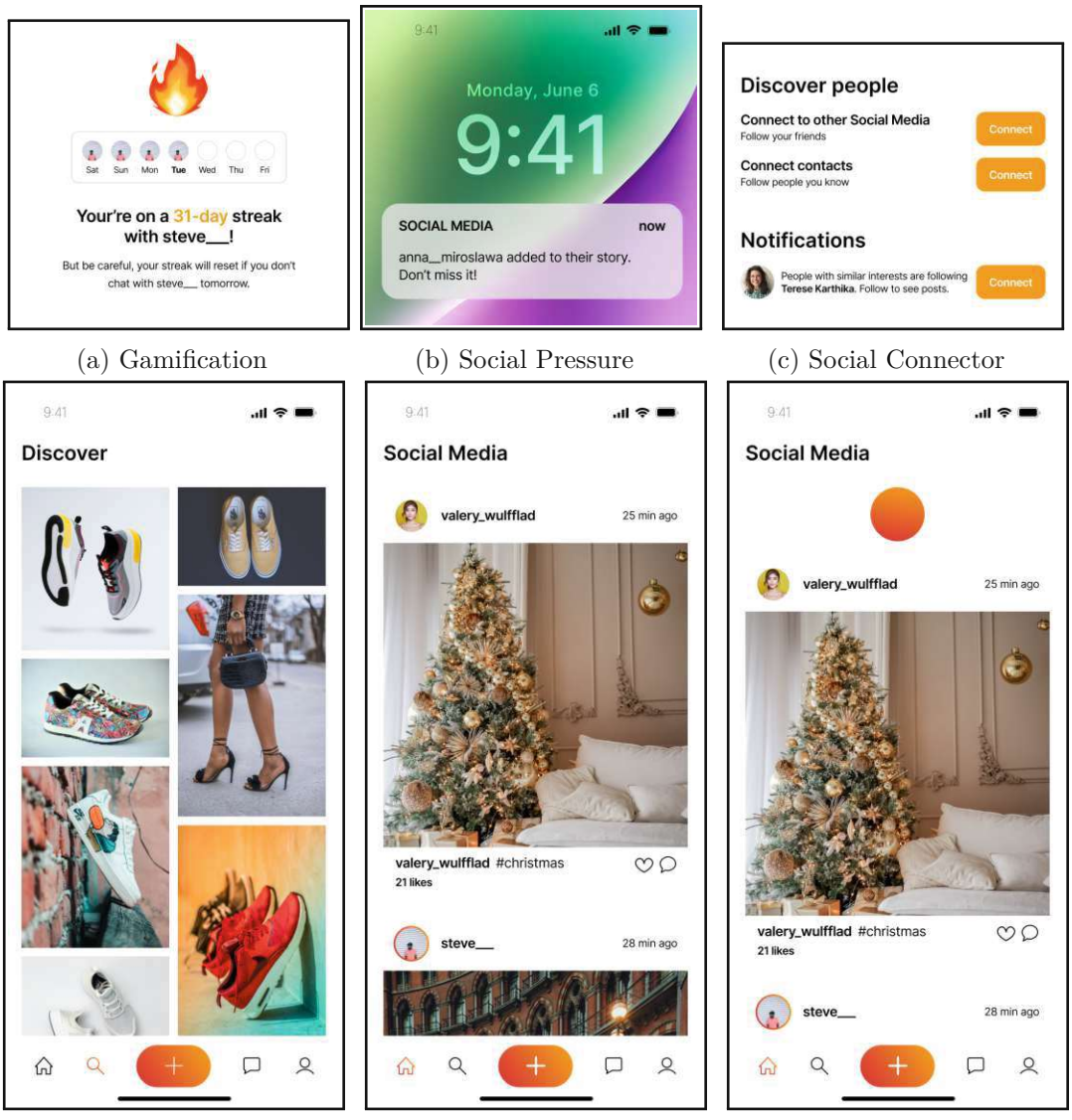
This dark pattern gets users hooked on the platform by providing customized content. The content can either be precisely accurate in a way that causes users to fall into a rabbit hole, or it can mismatch users' interests so that users need to search longer. In both ways, the content misaligns between the goals of the platform and the user's utility [56, 57]. In Figure 4.6d, the matching dark pattern mockup is presented, which displays exclusively images of shoes.

Endlessness

This dark pattern presents an infinite stream of content, which creates the illusion that new interesting content will continue indefinitely, while the content of the upcoming items is unpredictable. In some cases, content appears automatically without any user interaction, which removes the need for autonomous decision-making [56, 57]. Figure 4.6e shows a screenshot from the corresponding dark pattern video mockup, in which endless scrolling of different content is demonstrated.

Pull To Refresh

This dark pattern involves a mechanism where the user swipes (pulls) down on a mobile app to reload the page and load new content. This creates a sense of variable reward, generating excitement over whether new content will appear, similar to the rewards given by slot-machines [56, 57]. Figure 4.6f presents a screenshot of the dark pattern video mockup, demonstrating the pull to refresh mechanims.



(a) Gamification (b) Social Pressure (c) Social Connector

(d) Content Customization (e) Endlessness (f) Pull To Refresh

Figure 4.6: Social Engineering Dark Pattern Mockups

High-Level	Low-Level	Merged Patterns	Definition
Nagging	Nagging	Continuous Prompting [27]	Situations where users are annoyed by repeated attempts to get them to do something they did not intend to
Obstruction	Overcomplicated Process	Longer Than Necessary [27], Lacking Hierarchy [27], Too Many Options [27], Privacy Maze [27], Labyrinth [56]	Processes that are intentionally designed to be complicated by adding unnecessary steps, too many options (in inconsistent ways), or a lack of clear hierarchy
	Hindering Account Deletion	Clinging to accounts [56], Forced Grace Period [56]	Difficulties encountered by users while attempting to (immediately) delete their social media accounts
Sneaking	Sneaking Bad Default	Auto Accept Third Party Terms [56], Deceptive Snugness [27]	Instances where default selections are made by the platforms to serve their own interests, in worst case without users' awareness
	Expectation Result Mismatch	Misleading action [27], Plain Evil [56], Dead end [27], Decontextualising [27], Inconsistent Interface [27]	Situations where users expect a specific outcome but receive a different result, possibly due to exchanged functionality, dead ends, or information placed out of context
Interface Interference	False Hierarchy	Hidden in Plain Sight [27, 56], Look over there [27], Reduced Friction [56]	Designs that form the users' choice by enticing their attention away from undesired options to platform's preferred ones
	Trick Wording	Language discontinuity [27], Ambiguous wording or information [27], Conflicting information [27]	Inappropriate language is used to confuse users, leaving them left in misunderstanding
	Toying with Emotion	Emotional Steering [27], Persuasive Language [56]	Situations where emotions are used to steer the users' decisions in a particular direction
Forced Action	Forced Access	Forced Access Granting [56]	Instances where users are forced to give access to something to continue getting to their initial goal
Social Engineering	Gamification	Gamification [56], Social Investment [57]	Situations where rewards are used, pressuring users to continue using the platform to not lose their already achieved progress
	Social Pressure	Fear Of Missing Out [56], Regression Toward The Mean [56]	Instances where users are pressured to visit the platform to avoid missing out on important and popular topics
	Social Connector	Social Connector [56]	Situations where users are prompted to provide information about their friends or invite them to expand their network
	Content Customization	False Content Customisation [56], Recommendations [57]	Customizing content to manipulate users, either through falsified customization so that users have to search longer for rewards or through precise customization so that users fall into a rabbit hole
	Endlessness	Infinite Scrolling [56, 57], Autoplay [56, 57]	Endlessly and automatically loading content, eliminating the need for user decision-making or interaction
	Pull To Refresh	Pull To Refresh [56, 57]	Functionality that requires users to pull on the interface to reload the page

Table 4.1: Final Taxonomy of Dark Patterns in Social Media

4.1.7 Summary of Findings

The literature review of dark pattern taxonomies resulted in a new taxonomy of 15 low-level dark patterns categorized into six high-level dark patterns. In the next step, these 15 low-level dark patterns were transformed into image or video mockups.

4.2 Survey Results

This section presents the findings from the survey. First, general information on the organisation of the survey as well as demographics of the participants are presented, followed by the findings.

4.2.1 Organisation and Demographics

The survey was online from mid-January until mid-February of 2024. After the data processing, the final data set consisted of 143 participants, divided up into 126 social media users and 17 non-social media users. Out of the 126 social media users, 53 participants (42%) describe themselves as female, while 73 (58%) describe them as male. The range of age is broad, from the youngest being 19 years old to the oldest being 80 years old. Three quarters (75%) resided in Austria by the time they participated in the survey, 9% in the United States, 5% in Switzerland or Germany, and the other 11% elsewhere or did not answer. 87% of social media users consider themselves at least somewhat technically savvy, with 56% responding *Yes*, and 31% responding *A bit* to the question of whether they consider themselves technically savvy. The other 13% do not consider themselves technically savvy.

4.2.2 Findings: Problems with Social Media

The users' perceived problems were gathered in two different kinds of question formats, with different statements on a 7-point Likert scale and an open-ended question, as previously described in section 3.3.2. The *don't know* options on the Likert scale assessment were excluded from the following evaluation. Thus, each statement was answered by a different subset of the 126 social media user participants. The open-ended question asking for unfavorable features of social media was answered by 40 participants.

The users' ideas for changes were gathered with an open-ended text field, whereby 49 participants entered text into the text field for modifications or new features they would suggest if they could talk to social media platforms' managers, designers, and developers. Nine of them gave responses such as *"I do not know"*, or *"Nothing"*. The other 40 answers are divided up into concrete ideas and values.

The discovered problems are presented in the following sections, along with the concrete ideas for changes. Subsequently, the values that participants wished for are presented.

Social media is time-consuming and time-wasting

As shown in Figure 4.7, 60% of participants usually spend more time on social media than initially intended. 58% sometimes regret the time they spent on social media, and almost one-third of the participants are not satisfied with the amount of time they spend on social media. This is also evident in the open-ended responses, where participants expressed concern over the precious time lost while scrolling through social media.

Suggested solutions to reduce time spent: Participants expressed a desire for new features or platform adaptations to decrease their time spent. They suggested time limits, pagination features instead of endless scrolling, time bars, and regular reminders for breaks. One participant mentioned that social media platforms started hiding the mobile device's clock so that users would lose track of the time and they wished for these platforms to display the clock again.

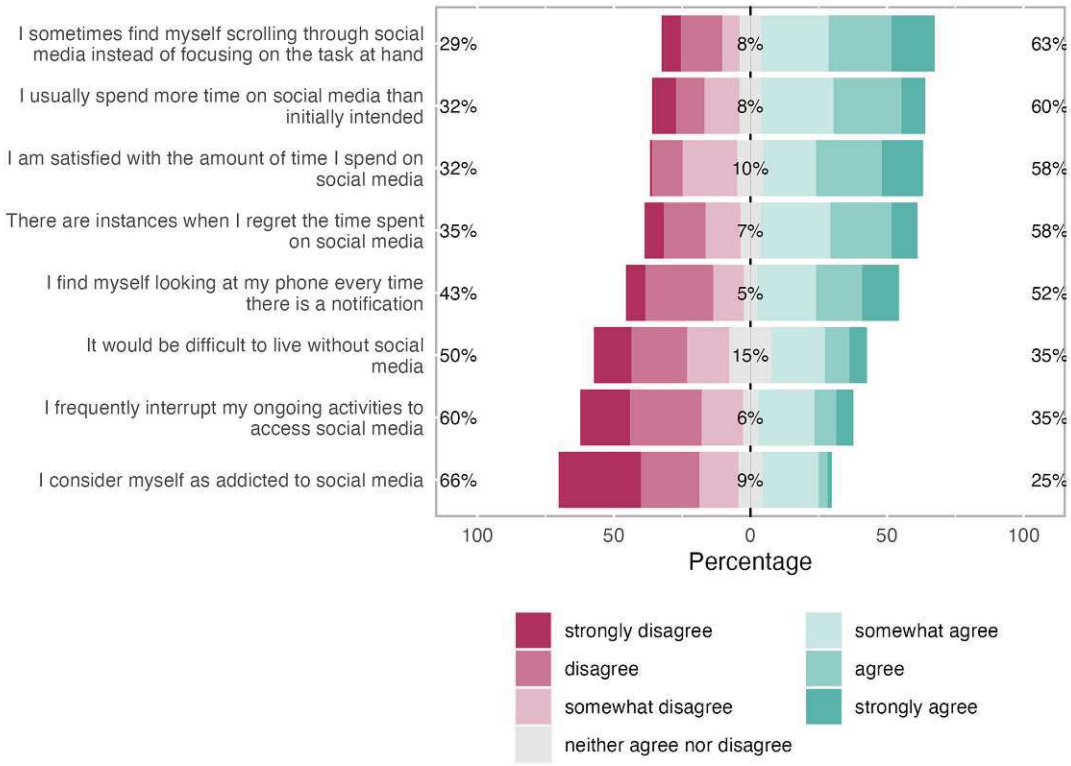


Figure 4.7: Survey Responses on Time, Focus, and Addiction Related Statements

Social media takes away users focus

Figure 4.7 also displays that one-third of the participants frequently interrupt ongoing activities to access social media, and more than half of participants find themselves looking at their phone every time there is a notification. In the open responses, one

person mentioned the misuse of notification indicators, signaling a received message when it's only that a friend of theirs has liked something, as annoying. Almost two-thirds of the participants sometimes find themselves scrolling through social media instead of focusing on the task at hand. Participants responded to the open-ended question that particularly Reels and TikTok bubbles lead to excessive usage and increased procrastination. The negative perceptions of these short-form videos were also reported due to their potential to capture attention and negatively impact attention spans. In general, social media is perceived as being able to *"take away your main focus"*.

Social media is addictive

One quarter agreed at least a little bit with considering themselves addicted to social media, and one-third believe it would be difficult to live without social media, as presented in Figure 4.7. Also, in the open-ended answers, some individuals used words like *addictive*, and one participant noted feeling pressured to use social media to avoid missing out on important events and news as social media is a central platform for particular communities. One person even highlighted the strategic tactics employed by social media platforms to extend user engagement. Another person said that social media is a *"source of endless information that is very easy to consume"* and that some of their associates spend lots of time on these platforms. Yet another participant said that excessive usage gets users down.

Suggested solutions to reduce time spent: Participants suggested removing addiction features and incorporating anti-addiction features.

Social media spreads fake news

Figure 4.8 illustrates that 83% of participants believe that social media spreads fake news. This is reflected in the open-ended responses, where participants noted disliking the spread of fake news. Two participants identified the clickbait function as a source that particularly intensifies this problem. Catchy and appealing headlines are chosen to create traffic, even though they involve false information. Since many people only read the headline, this results in the spread of fake news. Another reason mentioned was that social media does not sufficiently point out risks that lead to the random sharing of information. More than half of the participants indicated that they fact-check news articles or information they read on social media.

Suggested solutions to combat fake news: Users wished for more focus on the spread of fake news with better filters against fake information and fake profiles. While one person suggested launching a function allowing the community to edit clickbait titles to mitigate the spread of fake news, another one suggested implementing community notes like on Twitter.

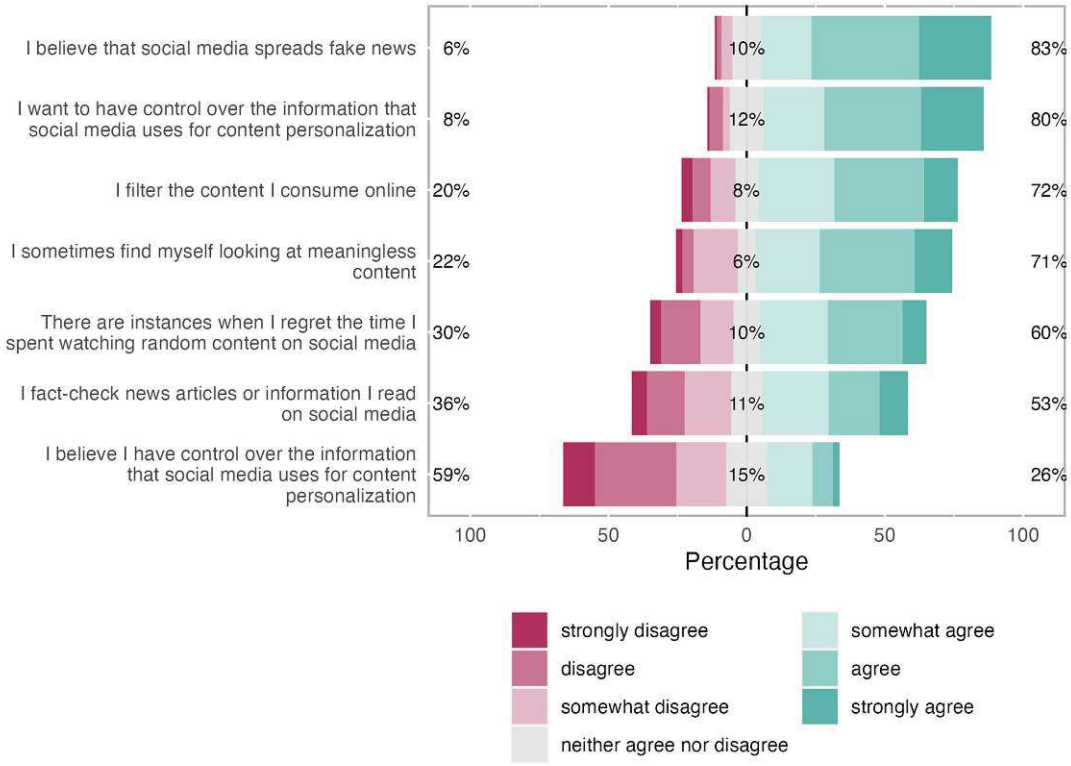


Figure 4.8: Survey Responses on Content-Related Statements

Social media does not provide meaningful content

As shown in Figure 4.8 59% of participants do not believe they have control over the information that social media uses for content personalization, and four out of five want to have control over the information that these platforms use for content personalization. The outcomes of content customization algorithms lead to a lack of eudaimonic happiness, which emphasizes meaningfulness and self-realization. In the open-ended responses, participants described the displayed content as *useless* and not creating an understanding of the topics they were interested in. Thus, they feel as if they could spend their spare time more usefully. Concerns were also raised about the inability to filter out undesirable or taboo topics, resulting in complete bans. 72% filter the content they consume online, and 71% sometimes find themselves looking at meaningless content, while 60% sometimes regret the time they spent watching random content on social media.

Suggested solutions to see desired content: Many participants criticized the content they see and wished for a different content customization algorithm, with one person even requesting, *"No algorithm"*. They want the ability to customize the algorithm by selecting preferences and content they're interested in, like only seeing longer videos instead of short ones, as well as a function to explicitly mark content as (not) helpful. One participant

stated a desire to have less personalized content, focusing on the exploration part rather than exploitation. Users only want to see posts from connected persons and no suggested profiles, and they request to see fewer or no ads. Additionally, users want to change not only the content shown to them but also the order of this content. They want the capability to edit the order of content or have a chronological order of content.

Social media is privacy invading

Half of the participants are concerned about their privacy and security on social media platforms, as shown in Figure 4.9. More than three-quarters do not believe that they have control over the personal information collected by social media companies about them, and almost two-thirds are concerned about the amount of personal information these companies possess about them. More than half of the participants perceive the content and ads they see on social media as an invasion of their privacy. This is echoed in the open-ended responses, where participants mentioned the perceived monitoring of their data as an unfavorable feature, explicitly mentioning instances where Google searches led to targeted advertisements for these products on social media. Another participant highlighted the uncertainty regarding the handling of private data.

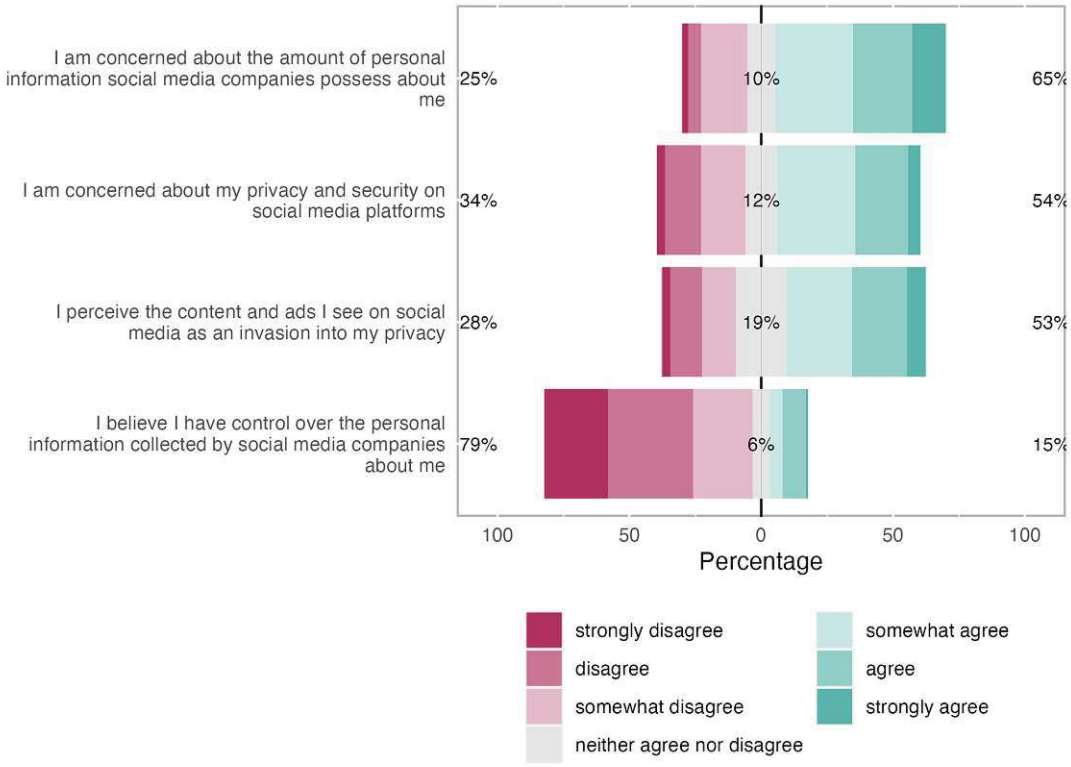


Figure 4.9: Survey Responses on Privacy-Related Statements

Social media is bad for some users' well-being

Figure 4.10 demonstrates that even though almost one-fourth of the survey participants agree that it has a positive impact on their mental well-being, almost one-third agree that social media has a negative impact on their mental well-being. In the open responses, one participant said that social media is *"not good for relaxation"*. Other drawbacks mentioned are hate messages, a depersonalization element, and social media leading to communication issues. Almost half of the participants agree that social media creates a sense of belonging and connection with others, whereas 14% feel lonely using social media.

Suggested solutions to increase well-being: One person wants mobbing to be banned.

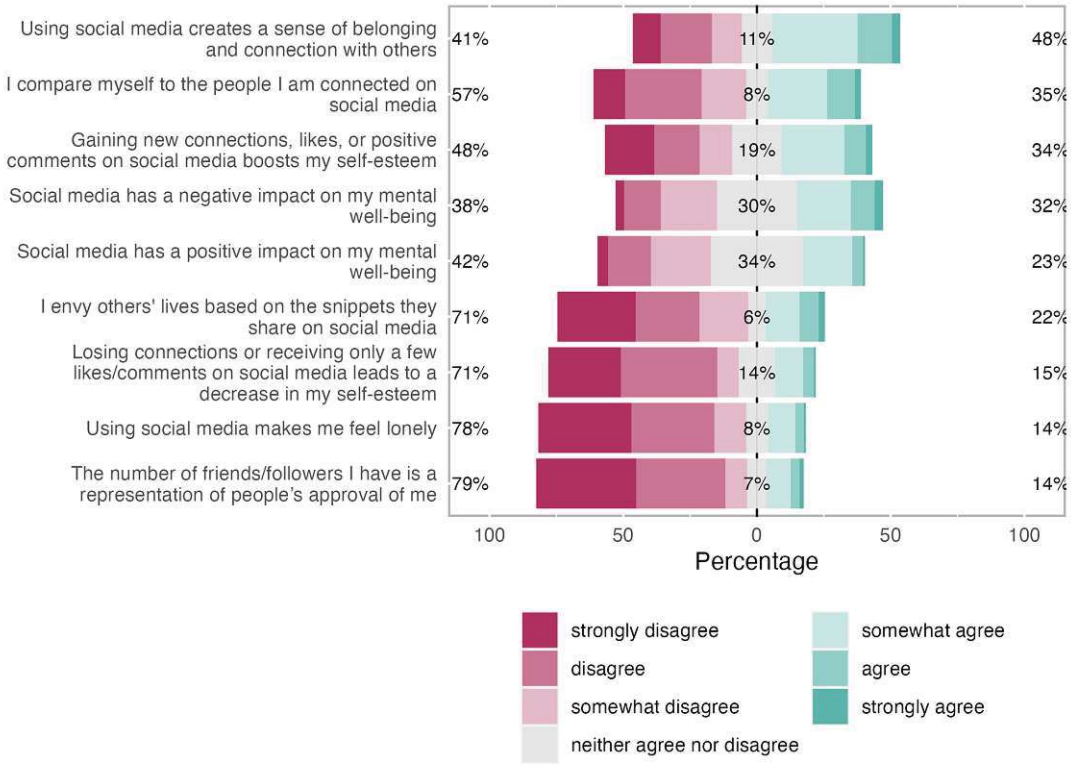


Figure 4.10: Survey Responses on Well-Being Related Statements

Social media fosters creation of unreal content and comparison among users

In the open-ended responses, concerns related to well-being centered around the creation of unreal content by social media users aiming to get as many followers as possible. Figure 4.10 illustrates that 14% of the participants agree with the statement that the number of friends/followers they have is a representation of people's approval of them. While one-third think that gaining new connections, likes, or positive comments on social

media boosts their self-esteem, 15% agree that losing connections or receiving only a few likes/comments on social media leads to a decrease in their self-esteem. However, this creation of unreal content not only results in fake news but also a distortion of reality. The tendency for individuals to post only the most positive aspects of their lives on social media leads to users comparing themselves to other users and a feeling of not doing enough. One-third compare themselves to the people they are connected with on social media, and one-fifth envy other's lives based on the snippets they share on social media. One person wrote as a response that *"bad influence on social media destroys one's life"*.

Suggested solutions against distortion of reality: Participants also mentioned a desire to hide the number of followers and likes on all posts as well as to remove the photo editing function. One person wants to require users to indicate whether they have edited a picture.

Other new features: Participants wished to see when another person viewed their profile and suggested launching forum-like features to facilitate discussions.

Values

Besides requesting both new features and the removal of existing ones, participants also requested better values. While one person asked for a general increase in ethics, other identified values include enhanced user autonomy, increased transparency, enhanced privacy, increased safety, and improved usability.

Enhanced User Autonomy: User autonomy emerged as a prominent theme in the responses, particularly concerning the amount of time they spend, with suggestions ranging from removing addiction features to incorporating anti-addiction features. Additionally, respondents desired more control over the content shown to them, for example, having the ability to select content topics of interest and order the results. One participant wanted more control over push notifications, another one over the duration of status visibility, and yet another one wanted the ability to skip YouTube ads. Furthermore, one respondent wanted everyone to have access to all platform functions, meaning that the availability of functions does not depend on the number of followers one person has.

Increased Transparency: Surveyed users expressed a desire for increased transparency regarding ads, AI, data control, and personal content handling. One person suggested adding comments like, "You see this because of this". Another person described the degree of transparency they wish for, stating, *"I want to have important information precisely in one sentence about why the account will be deleted only after 30 days, and not 15 lines about what happens."*
In addition to requesting increased transparency concerning system actions, participants also expressed desire for transparency regarding the actions of other social media users. They wanted peer users to be transparent by proposing a requirement for users to declare whether a photo has been edited or not.

Enhanced Privacy: Participants wanted less tracking of information since they sometimes feel spied on. Additionally, they opposed personal data transfer, with one person request-

ing to ban it. Another person asked about the possibility of removing a mention/tag by another person.

Increased Safety: Surveyed users highlighted the need for better handling of fake information, fake profiles, and hate postings. In addition, they wanted more content monitoring and censorship and asked for a ban on mobbing. One person stated that social media is not a legally free area.

Improved Usability: Users desired increased usability by reducing functionality and focusing on ease of use. For instance, they wish for easier processes for editing their settings or deleting their account.

4.2.3 Findings: Participants attempts to combat social media problems

The following attempts to combat social media problems could be found in the data from single and multiple-choice questions about social media habits.

Attempts to reduce time: When asking participants whether they have ever tried to reduce the amount of time they spend on social media, 54 participants, which equals 43%, answered yes. Among this subset of participants, 87% were able to reduce their time by at least a little bit.

Looking at only participants who are concerned about overspending time on social media, it becomes clear that they are willing to make changes as they have already made some effort against it. In this case, concerned participants are defined by agreeing with the statement *I usually spend more time on social media than initially intended* and disagreeing with *I am satisfied with the amount of time I spend on social media.* Out of a total of 37 concerned overspenders, 59% have used time-tracking apps or features to control their daily social media usage. Furthermore, 65% of them have tried to reduce the amount of time they spend on social media.

Attempts to decelerate the spread of fake news: Half of the 105 participants (53%) who believe that social media spreads fake news, fact-check news articles or information they read on social media.

4.2.4 Findings: Perception of Dark Patterns

Table 4.2 presents an overview of how participants evaluated the individual dark pattern mockups on the UEEQ. It outlines the overall mean and the item means per pattern. A threshold of 0.8 was used.

The table presents the following information:

Nagging received a negative evaluation from surveyed users, averaging -0.92, making it the second most negative overall mean among all dark pattern evaluations. Participants reported perceiving it as *pressuring*, *obstructive*, *annoying*, *not interesting*, *unfriendly*, *deceptive*, and *boring*. On the other hand, they perceived it as *non-addictive*. The

Pattern	Overall Mean	unfriendly vs friendly	annoying vs enjoyable	obstructive vs supportive	unpredictable vs predictable	cluttered vs organized	inefficient vs efficient
NAG	-0.92	-1.73	-1.97	-2.08	0.16	-0.71	-0.71
OP	-0.49	-0.88	-0.57	-0.35	-0.29	-1.20	-0.42
HAD	0.02	-0.47	-0.44	-0.22	0.70	0.40	0.43
SBD	0.35	-0.65	-0.87	-0.11	0.71	0.95	0.34
ERM	-0.58	-1.36	-1.00	-0.72	0.03	-1.19	-0.34
FH	-0.86	-2.13	-1.97	-1.17	0.46	0.38	-0.25
TW	-0.78	-1.82	-1.85	-1.35	-0.91	0.52	-0.23
TWE	-0.85	-2.15	-2.15	-1.85	0.21	0.69	0.03
FA	-0.93	-1.44	-1.69	-1.23	-0.95	-0.15	-0.82
GAM	-0.31	-0.87	-0.73	-0.55	0.44	-0.10	0.34
SP	0.00	-0.60	-0.67	-0.29	0.61	0.72	0.41
SC	-0.02	-0.71	-0.73	0.36	0.03	0.90	0.57
CC	0.08	-0.15	-0.06	0.17	0.42	-0.08	0.03
END	0.01	-0.10	0.05	-0.20	-0.50	0.22	0.81
PTR	0.48	0.80	0.34	1.07	-0.28	0.75	0.97

Pattern	addictive vs non-addictive	pressuring vs suggesting	covert vs revealed	deceptive vs benevolent	confusing vs clear	complicated vs easy	not interesting vs interesting	boring vs exciting
NAG	0.89	-2.26	0.78	-1.58	-0.23	0.41	-1.79	-1.26
OP	0.63	-0.39	0.18	-0.61	-1.12	-0.83	-0.20	-0.48
HAD	-0.30	-0.46	0.56	-0.76	0.53	0.43	0.03	-0.15
SBD	1.26	-0.33	1.49	-0.56	1.30	0.98	0.28	-0.09
ERM	0.88	-0.65	0.06	-1.28	-1.29	-0.59	-0.24	-0.19
FH	0.86	-2.11	-0.43	-2.09	-0.76	0.44	-0.76	-0.86
TW	1.07	-1.29	-0.15	-1.76	-0.86	-0.66	-0.06	-0.44
TWE	-0.29	-1.92	0.25	-1.97	-0.40	0.53	-0.83	0.00
FA	0.29	-1.59	-0.44	-1.52	-0.90	-0.79	-0.75	-0.58
GAM	-0.88	-1.83	0.62	-0.83	0.61	1.06	-0.63	-0.23
SP	-0.28	-1.02	0.19	-0.31	1.16	1.72	-0.86	-0.40
SC	0.21	-1.12	0.38	-0.89	0.44	0.97	-0.22	-0.18
CC	0.17	-0.71	0.87	-0.61	0.58	1.09	-0.16	-0.35
END	-1.47	-0.48	0.19	-0.82	0.49	1.46	0.18	0.35
PTR	-1	-0.04	1.11	0.07	0.59	1.66	0.62	0.17

Table 4.2: Overall Mean and Item Means per Dark Pattern. Item Means $\leqslant$ -0.80 are Highlighted in Red, and Item Means $\geqslant$ 0.80 are Highlighted in Green.
NAG = Nagging, OP = Overcomplicated Process, HAD = Hindering Account Deletion, SBD = Sneaking Bad Default, ERM = Expectation Result Mismatch, FH = False Hierarchy, TW = Trick Wording, TWE = Toying with Emotion, FA = Forced Access, GAM = Gamification, SP = Social Pressure, SC = Social Connector, CC = Content Customization, END = Endlessness, PTR = Pull To Refresh

expectations before the survey that it would be perceived as annoying and pressuring were confirmed.

The mockup of the *Overcomplicated Process* dark pattern instance scored an overall neutral mean of -0.49. Participants evaluated it as *confusing*, *cluttered*, *unfriendly*, and *complicated*. Besides the negative evaluations, the pattern did not receive any positive evaluations. Rather, users had no strong opinions on this dark pattern, with most means falling between opposing characteristics. The pre-survey assumptions regarding its perception as cluttered and confusing were confirmed.

Hindering Account Deletion was evaluated neutrally, without any spikes or strong opinions on a particular adjective. This resulted in an overall mean of 0.02. Prior to the survey, it was assumed that this dark pattern instance would be perceived as deceptive and obstructive. The results show that the surveyed users do not perceive it as either of them, contradicting my assumptions. This could be due to the fact that some people perceive a grace period as a helpful feature to prevent impulse actions.

The *Sneaking Bad Default* dark pattern instance scored an overall mean of 0.35. Participants positively rated it as *revealed*, *clear*, *non-addictive*, *easy*, and *organized*. However, on the other hand, they rated it as *annoying*. The expectations that it would be perceived as annoying and covert were partially confirmed. While it was perceived as annoying, it was not covert at all, but rather the opposite–revealed.

Expectation Result Mismatch averaged an overall mean of -0.58. Survey participants evaluated it as *non-addictive* but also as *unfriendly*, *confusing*, *deceptive*, *cluttered*, and *annoying*. The evaluation of the pre-survey assumptions varies, with deceptive and confusing being confirmed and unpredictable, obstructive, and complicated not being confirmed.

Besides a positive evaluation of *False Hierarchy* being *non-addictive*, participants generally viewed it neutrally or negatively. This dark pattern instance achieved an overall negative mean of -0.86, placing it as the third worst overall mean among all dark pattern evaluations. Participants assessed it as *unfriendly*, *pressuring*, *deceptive*, *annoying*, *obstructive*, and *boring*. One of the pre-survey assumptions that this dark pattern instance would be perceived as obstructive was verified, while the other expectation that it is perceived as covert was not confirmed.

The *Trick Wording* pattern instance attained a mean of -0.78, ranking it among the top five worst overall means. Surveyed users rated it as *annoying*, *unfriendly*, *deceptive*, *obstructive*, *pressuring*, *unpredictable*, and *confusing*. In contrast, it was assessed as *non-addictive*. While the assumption about this instance being covert was not confirmed, the other assumptions about it being deceptive, unpredictable, and confusing were verified.

Toying with Emotion scored an overall negative mean of -0.85, placing it fourth among the worst-rated mean of all dark patterns. When compared to other dark patterns, *Toying with Emotion* was assessed as the most *annoying* and most *unfriendly*. Additionally, it

was perceived as *deceptive*, *pressuring*, *obstructive*, and *not interesting*. Both pre-survey assumptions regarding it being perceived as unfriendly and pressuring were validated.

The *Forced Access* dark pattern received the worst overall mean from the participants, at -0.93. It was perceived as *annoying*, *pressuring*, *deceptive*, *unfriendly*, *obstructive*, *unpredictable*, *confusing*, and *inefficient*. As anticipated before the survey, this dark pattern was perceived as pressuring.

The mockup of the *Gamification* dark pattern averaged an overall mean of -0.31. While participants rated it negatively as *pressuring*, *addictive*, *unfriendly*, and *deceptive*, they positively evaluated it as *easy*. The anticipation of this dark pattern being perceived as pressuring and addictive was confirmed.

Social Pressure received a neutral evaluation, with an overall mean of 0.00. While it was positively assessed as *easy* and *clear*, this was counterbalanced by users' perception of it as *pressuring* and *not interesting*. Before the study, three different assumptions were made, one of which was validated: this dark pattern is perceived as pressuring. However, it was neither seen as addictive nor annoying, not confirming the other two assumptions.

The *Social Connector* dark pattern was perceived neutrally, with an overall mean of -0.02. Participants rated it as *pressuring* and *deceptive*, while also assessing it as *easy* and *organized*. The pre-survey assumption regarding this dark pattern being perceived as pressuring was validated.

Content Customization was evaluated neutrally, with an overall mean of 0.08. It was perceived as *easy* and *revealed*. Aside from that, there were no other significant associations with this feature. The two pre-survey assumptions regarding users' perception of it as enjoyable and addictive were not confirmed.

The *Endlessness* dark pattern instance averaged an overall neutral mean of 0.01. Survey participants perceived it as *easy* and *efficient* but also *addictive*. Before the survey, this dark pattern was expected to be perceived as addictive, enjoyable, and supportive. The former expectation was validated, whereas the latter two were not confirmed.

Pull To Refresh achieved the best overall mean, at 0.48. While participants indicated perceiving it as *addictive*, they also rated it positively for being *easy*, *revealed*, *supportive*, *efficient*, and *friendly*. It did not confirm the pre-survey assumption of being enjoyable. However, it did confirm the other expectation of being perceived as addictive.

The following sections present five findings from the dark pattern evaluation.

1. Similar Evaluations Among Dark Patterns within Categories

When comparing dark patterns of one high-level category at a time, there are some similarities in the evaluations by the participants. This is depicted in the correlation coefficients in Table 4.3, as well as in the graph in Figure 4.11. Correlations were found to exist between low-level dark patterns within all high-level dark patterns, except for the low-level dark patterns in *Obstruction*. For instance, the three patterns of *Interface*

Interference are consistently rated as *pressuring, deceptive, obstructive, annoying,* and *unfriendly*, and score an overall correlation of 0.77. Another pattern emerges under the *Social Engineering* category, scoring a correlation of 0.72, where all patterns were evaluated as *easy* rather than *complicated*. In addition, they are generally perceived as *clear* rather than *confusing*, with users' tendency to evaluate them as *efficient*.

High-Level Dark Pattern	Correlation
Nagging	-
Obstruction	-0.23
Sneaking	0.29
Interface Interference	0.77
Forced Action	-
Social Engineering	0.72

Table 4.3: Correlation Coefficients between Patterns in each High-Level Dark Pattern

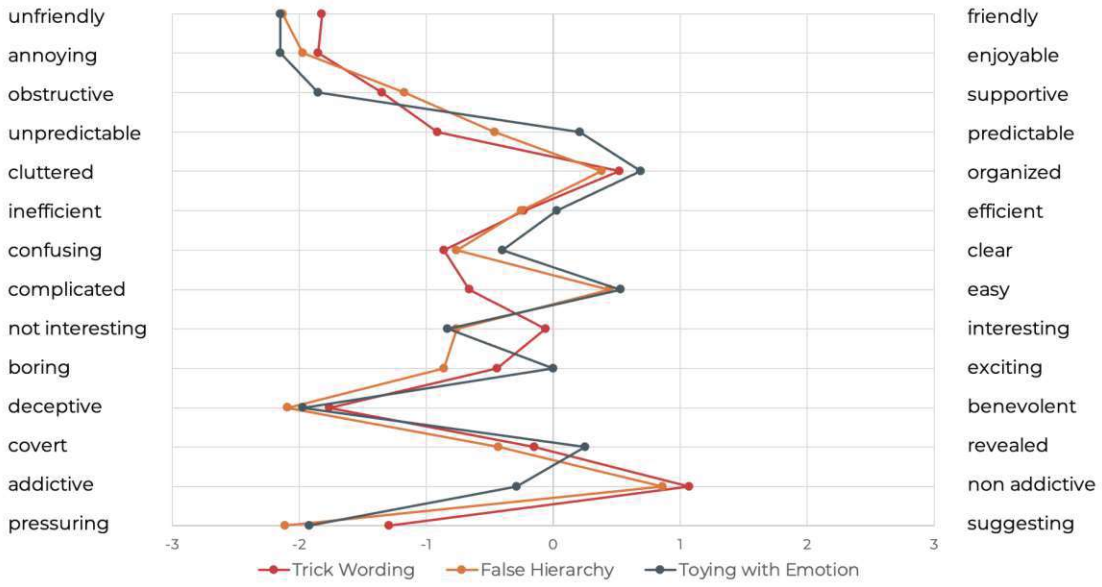


Figure 4.11: Interface Interference Dark Patterns scored Similar Evaluations

2. Deceptive Dark Patterns are Not Attractive

When taking a closer look at the various UEEQ scale scores, one can observe a correlation of 0.93 between the *Attractiveness* scale and the *Deception* score. In other words, dark patterns perceived as *deceptive* are also perceived as *not attractive*. Taking a look at Figure 4.12, dark patterns generally do not receive positive assessments on the *Attractiveness* scale, except for the *Pull To Refresh* dark pattern instance. This particular instance has a slightly positive value of 0.58. The same is true for the deceptive scores:

Pull To Refresh is the only dark pattern perceived as *benevolent* instead of *deceptive*. Every other pattern was evaluated negatively.

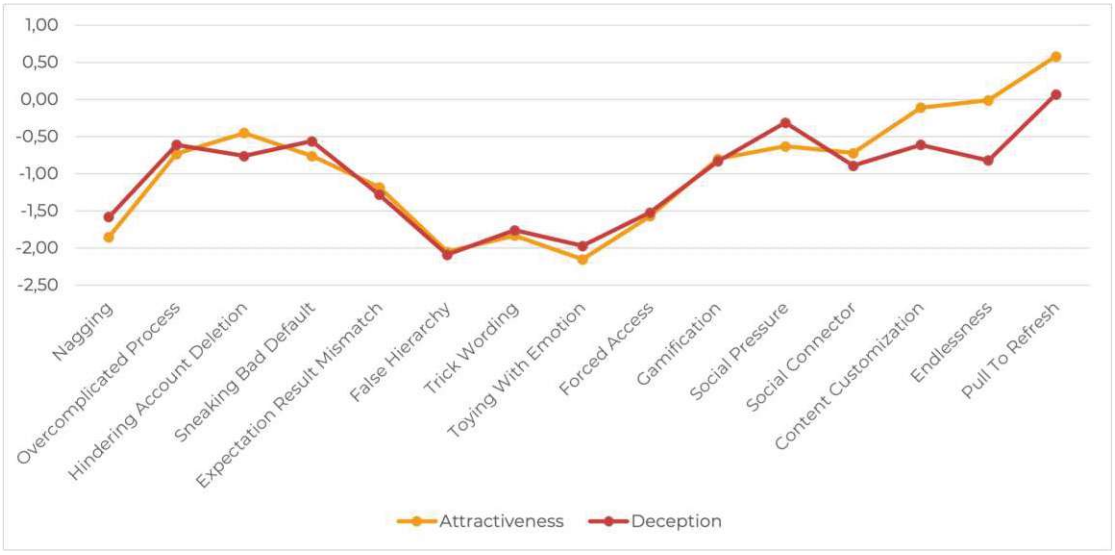


Figure 4.12: Correlation between Deception and Attractiveness

3. Dark Patterns are Pressuring, Deceptive, Annoying, and Unfriendly

When comparing the different character pairs used in the questionnaire and calculating their average evaluation among all dark patterns, *pressuring* has the lowest mean of -1.08, followed by *deceptive* at -1.03, *annoying* at -0.95, and *unfriendly* at -0.95. This is depicted in Figure 4.13.

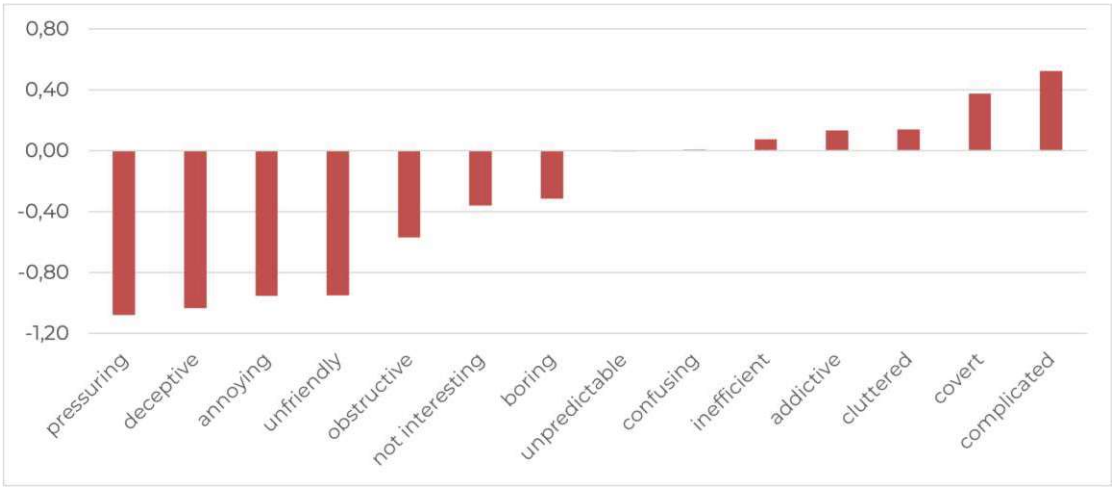


Figure 4.13: Average Evaluation of All Dark Patterns per UEEQ Item

4. Forced Access was Perceived Worst among All Dark Patterns

When ranking dark patterns by their overall mean, *Forced Access* has the worst overall mean. It is followed closely by *Nagging*, *False Hierarchy*, *Toying with Emotion*, and *Trick Wording*.

Upon closer examination of the manipulative items scores *Pressure, Addiction, Covert*, and *Deception*, the dark pattern *Forced Access* also ranks among the worst five dark patterns in three out of these four scales. Only *Toying with Emotion* and *False Hierarchy* share the same number of appearances as being three times within the worst 5. Then, *Nagging*, *Trick Wording*, and *Gamification* appear twice under the worst 5. *Overcomplicated Process*, *Hindering Account Deletion*, *Expectation Result Mismatch*, *Endlessness*, and *Pull To Refresh* each appear once under the worst 5.
The dark pattern instances *Sneaking Bad Default*, *Social Pressure*, *Social Connector*, and *Content Customization* are not under the worst 5 in any category.

Here are the worst five in each of the new UEEQ categories and the Attractiveness scale:

- Bottom 5 Covert: (1) Forced Access, (2) False Hierarchy, (3) Trick Wording, (4) Expectation Result Mismatch, (5) Overcomplicated Process
- Bottom 5 Pressure: (1) Nagging, (2) False Hierarchy, (3) Toying with Emotion, (4) Gamification, (5) Forced Access
- Bottom 5 Deception: (1) False Hierarchy, (2) Toying with Emotion, (3) Trick Wording, (4) Nagging, (5) Forced Access
- Bottom 5 Addiction: (1) Endlessness, (2) Pull To Refresh, (3) Gamification, (4) Hindering Account Deletion, (5) Toying with Emotion
- Bottom 5 Attractiveness: (1) Toying with Emotion, (2) False Hierarchy, (3) Nagging, (4) Trick Wording, (5) Forced Access

5. Dark Patterns with Good and Bad Evaluations

Even though the dark pattern instances were generally assessed as negative or neutral, there are some that got positive evaluations in single categories that I want to highlight here:

- **Social Pressure**: is pressuring (-1.02) but also easy (1,72) and clear (1,16).
- **Endlessness**: is addictive (-1.47) but easy (1.46).
- **Pull To Refresh**: is addictive (-1.00) but also easy (1.66) and supportive (1.07).
- **Gamification**: is pressuring (-1.83) but also easy (1.7)

4.2.5 Summary of Key Findings

1. Eight social media problems were identified: Social media is time-consuming and time-wasting, social media takes away users focus, social media is addictive, social media spreads fake news, social media does not provide meaningful content, social media is privacy invading, social media is bad for some users' well-being, and social media fosters creation of unreal content and comparison among users.

2. Users have made attempts to reduce time spent and decelerate the spread of fake news.

3. Some of the pre-survey assumptions of how participants would rate particular dark patterns were verified, for example, *Nagging*, while others were not, for example, *Hindering Account Deletion*.

4. Dark patterns within a high-level dark pattern category were perceived similarly.

5. Deceptive dark patterns are not attractive, and dark patterns, in general, are *pressuring, deceptive, annoying*, and *unfriendly*.

6. *Forced Access* performed worst in the evaluation of the overall mean and is perceived as most *covert*.

7. Some dark patterns were not exclusively evaluated negatively on the UEEQ but also positively in some aspects.

While the last five listed findings answer part two of Research Question 1 (section 1.2), the first two listed findings do not answer any particular research question. Rather, they inspired further investigation and uncovered where bright patterns are needed, thus providing a basis for Research Question 2 (section 1.2).

CHAPTER 5

Findings from Exploring Bright Patterns

This chapter presents the findings of the exploratory bright pattern research, which involved inverting dark patterns into bright patterns, resulting in three newly designed bright patterns that were subsequently assessed:

- *Nagging* to reduce time spent
- *False Hierarchy* to decelerate the spread of fake news
- *Toying with Emotion* to influence the content algorithm

5.1 Designed Bright Patterns

After gathering first ideas, the big topic was narrowed down to a selection of three dark pattern types. The three least attractive dark patterns were chosen, namely *Nagging*, *False Hierarchy*, and *Toying with Emotion*. They were repurposed to work against three social media problems, aiming at reducing time spent, viewing more meaningful content, and decelerating the spread of fake news.

The following sections describe the three dark pattern techniques theoretically, provide examples, present the dark patterns created in the first part of this thesis, and explain how they were incorporated into the new proposed bright patterns.

5.1.1 Nagging

This pattern was not a result of merging different patterns, as this technique was only identified once in the three literature sources. It represents the *Continuous Prompting* pattern presented in the guidelines by the European Data Protection Board [27].

The technique utilizes annoying distractions to push users into accepting or confirming the platforms' intentions. If users do not confirm or accept, the platform repeatedly prompts them, disrupting the convenient use of the service. In some cases, the platform also provides explanations of why the user should accept it. This pattern makes users annoyed and creates a fear of having to refuse the request each time in order to use the platform [27]. Thus, they are time and nerve-consuming and won't disappear until users give in and accept the provider's offer/suggestion/request [14, 27, 36]. Examples of this dark pattern strategy and type are pop-ups or audio notices [36]. Phrases like *"Maybe later"* and *"Not now"* are used so that by selecting this option, the user implies a potential commitment to do it later on. This enhances the probability that the user will feel compelled to eventually do what the technology wants them to do [33].

The mindmap attached in the appendix as Figure 7.2 is a collection of the different kinds of *Nagging* examples found, each illustrated with one example. Most of the examples found require users to take immediate action in order to continue what they were doing, for example with a window popup or as a step in a process. However, there are also smaller elements that are overlayed on the user interface, which do not necessarily need user input in order to continue fulfilling the task at hand. They serve more as annoying distractions. Another form of *Nagging* is a setting option that, upon selection, will only be active for a specific amount of time and needs to be reselected after that. Yet another example found is the utilization of a red notification point in the settings, which will keep showing up until the user changes the settings to the company's desired options, for example, until the user finishes setting up their iPhone.

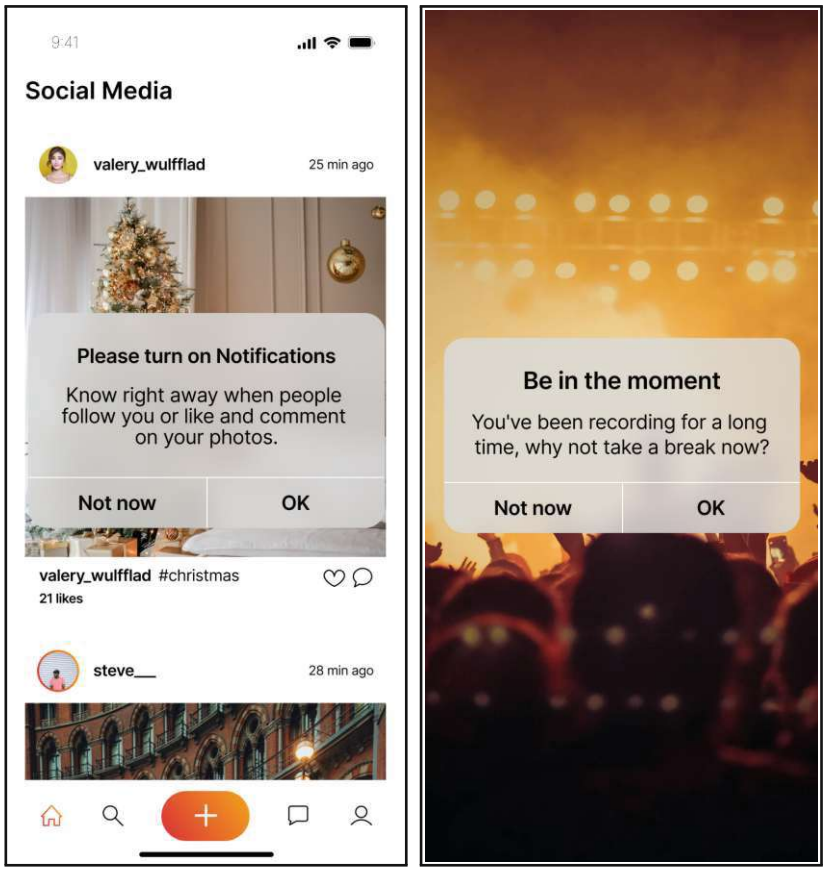
Dark Pattern to get Users to Accept Notifications

The dark pattern mockup designed in the first part of this thesis was a reproduction of examples found in literature [14, 36]. The mockup shown to the survey participants was a video demonstrating the frequency of the popups display, Figure 5.1a provides only a screenshot of the popup. The intention of this pattern is to manipulate users into accepting to turn on notifications. There is a popup asking the user to turn on their notification, but there is no button to decline it. Thus, the user is only able to accept it or tab on *"Not now"*, which will ask the user over and over again until they just accept it.

New Bright Pattern to Reduce Time Spent

The *Nagging* technique was used to motivate users to spend less time online and more time offline. Similar to the dark pattern mockup, the bright pattern mockup of the *Nagging* technique was also a video mockup, and Figure 5.1b presents only a screenshot of it. The video mockup shows a smartphone interface recording a sequence of a concert and frequently a popup was displayed as an overlay prompting the user to take an action. The popup's text states in bold font *"Be in the moment"* and underneath is a capture *"You've been recording for a long time, why not take a break now?"* and there were two possible options, namely *"Not now"* and *"OK"*. While the popup was displayed, the video

did not stop in the background. The frequency of the displaying of the popup should annoy users until they click on *"OK"*.



(a) Dark Pattern (b) Bright Pattern

Figure 5.1: Nagging Mockups

5.1.2 False Hierarchy

This pattern combines the patterns *Hidden in Plain Sight* [27, 56], *Look over there* [27], and *Reduced Friction* [56].

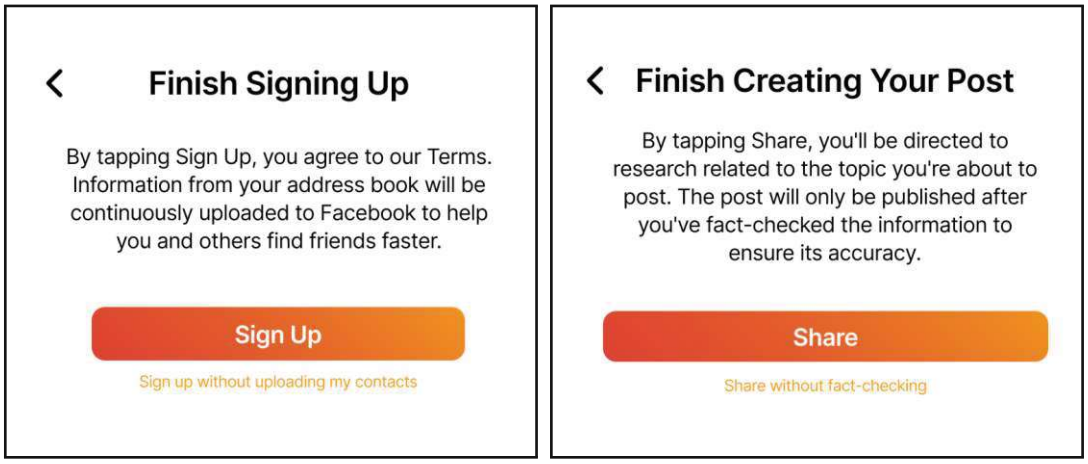
This technique makes use of style or other techniques to manipulate users' actions. Crucial information or options that do not align with business interests are styled less attention-grabbing and less accessible than the ones that align with the platform's preferences [27, 56]. Options that are equal and, thus, should be equally designed, are differently designed so that the user thinks one option is the best or even the only possible option [36]. In particular, the patterns utilize small font sizes or colors with too little contrast with the background to impede readability, and it is mostly applied to links and buttons [27]. This way, the user's attention is drawn to the options or actions that

are desired by the platforms [27, 56], making spotting crucial information cumbersome and time-consuming [27]. When it comes to presenting different user options, this dark pattern is often combined with *Sneaking Bad Default* [27].

Figure 7.3 in the appendix presents the mindmap of the different *False Hierarchy* techniques that were found. Most of the examples presented two equal options that were unequally designed to encourage the user to select the desired option. Some examples even looked as if the desired option was mandatory. Moreover, by employing *False Hierarchy*, some examples hide functions, important information, or options by placing them out of context. Furthermore, there are some design manipulations where a multiple-choice selection looks like a single-choice selection, or the user gets distracted by sound.

Dark Pattern to get Users to Upload their Contacts

Figure 5.2a shows the dark pattern mockup that was designed in the first part of this thesis based on an example found in literature [54]. The intention of this pattern is to manipulate users to upload their contacts in order to get more user data. During the account setup, there are two options, namely sign up with or without uploading their contacts. Even though these options are equal, the option that is desired by the company looks like a big button and only says *"Sign Up"*, while the other option's font has low contrast and looks more like additional information. Moreover, the button text does not say *"with uploading my contacts"* causing users who do not read the information in between the headline and the button to be unaware of what they are consenting to.



(a) Dark Pattern (b) Bright Pattern

Figure 5.2: False Hierarchy Mockups

New Bright Pattern to Decelerate the Spread of Fake News

The *False Hierarchy* technique was used to motivate users to fact-check the content they are about to post online. The finished pattern, presented in Figure 5.2b, serves as a last

step in the post-creation process. It says in bold font *"Finish Creating Your Post"* and has one big button with *"Share"* and small low-contrast text beneath it stating *"Share without fact-checking"*. In between the heading and the button, there is a text explaining that once the user clicks on the Share button, they will be directed to research related to the topic they are about to post and that their post will only be published after they have fact-checked the information. The intention of the two options being so differently designed should encourage users to click on the *"Share"* option.

5.1.3 Toying with Emotion

This pattern is a combination of *Emotional Steering* [27] and *Persuasive Language* [56].

Toying with Emotion intentionally creates emotions in users in order to push their decision into a particular direction [27, 56]. While the pattern description of *Persuasive Language* by Mildner et al. [56] focuses on the utilization of language to create these emotions, the description of *Emotional Steering* in the guidelines by the EDPB [27] includes the utilization of wordings, visuals, images or others to trigger emotional reactions in the users.

Brignull sets focus on uncomfortable emotions, such as guilt or shame, and calls the corresponding dark pattern *Confirmshaming* [14], which can be seen as a subcategory of *Toying with Emotion*. This dark pattern is often used in an opt-out context, where the words are phrased so that the user feels bad about not going with the suggested option [14]. For example, a declining option is *"No thanks, I hate saving money"* [52]. The gaming and lotteries website Iwin.com displays *"What's the deal? You don't call, you don't log in... Is it me?"* to get an unlogged user to log into the system [33]. According to Fogg, language is never neutral, not even used by a computing system [33]. This dark pattern makes usage of the framing effect cognitive bias by structuring and expressing undesired options negatively. The framing effect states that users make different decisions from the same information based on how it is presented to them [52]. To a certain extent, these dark pattern instances are designed to closely mimic human behavior, aiming to evoke a sense of responsibility within the user to comply with the system's suggestions [6]. Small changes to the content are made to create a social bond between the user and the software, for example, welcoming them by their name [33].

Toying with Emotion in this thesis serves as an umbrella term for all dark patterns aiming to create an emotion. Some make users feel good, safe, or rewarded by creating a highly positive outlook, and others make users feel anxious, guilty, or punished by a highly negative outlook [27]. This dark pattern can be found throughout the user journey. However, its utilization to push users to share more data is, according to the EDPB, especially strong in the sign-up process since users want to use the platform as soon as possible. Thus, they are in a rush to get done with the process in order to finally use the application, leading to not reflecting on their decisions and being more likely to fall for the deceptive techniques utilized [27]. The EDPB state that vulnerable users, including children and elderly people, might be more impacted by this pattern since they do not

understand processing activities. The authors appeal to social media platforms to use appropriate language, including tone and style, especially when social media services are specifically targeted to these vulnerable people [27].

A subset of the mindmap of the found *Toying with Emotion* examples is depicted in the appendix in Figure 7.4. Most of the found examples showed two options and the option against the business goal was emotionally blaming the user for not caring about themselves or their pets, for being too risky, for being a bad person, or for multiple other reasons. Sometimes, both options or options aligning with the business goal were emotional, and yet other examples had an emotional question. Other examples had several or only one option to blame the users. Moreover, some examples utilize emotional images, long emotional texts, or images and texts to manipulate users. Impersonation or making users feel confident in being themselves can also lead to *Toying with Emotion*. While some examples were rather mild, some were really strong, accusing the user of being ignorant and complicit in hate.

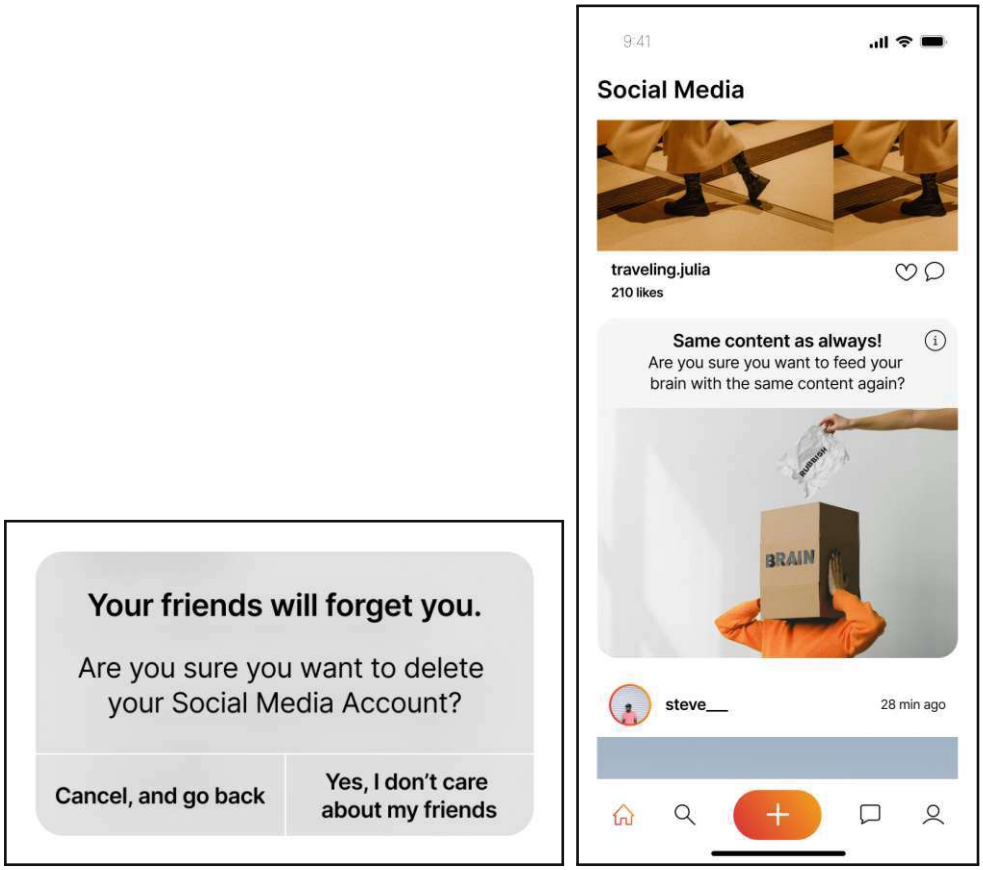
Dark Pattern to Prevent Users from Deleting their Accounts

The created dark pattern, depicted in Figure 5.3a, is a combination of examples found in literature [27] and on websites[1]. Its intention is to create a feeling of guilt, thus manipulating users into not deleting their social media accounts. In bold font, it says *"Your friends will forget you"*, followed by the question *"Are you sure you want to delete your Social Media Account?"* and then there are two options, namely *"Cancel, and go back"* and *"Yes, I don't care about my friends"*. The intention of these mean phrases was for users not to delete the account since they need to acknowledge their ignorance by tapping on this option.

New Bright Pattern to Influence the Content Algorithm

The *Toying with Emotion* technique was used to motivate users to explore content that was different from usual content. The underlying idea is that feeding the content algorithm with diverse posts will result in displaying a variety of posts. Thus, hopefully, this will decrease the feeling of watching useless content. Moreover, it would reduce the distortion of reality which is created by seeing the same things all the time. The pattern was designed similarly to a social media post and was shown in the feed. It showed a picture where someone throws a paper with the label *"RUBBISH"* into someone's brain, illustrated by a cardboard box. Above the image there was a text saying in bold *"Same content as always!"*, and then underneath *"Are you sure you want to feed your brain with the same content again?"*. This creation of guilt should manipulate users to look for different content. The finished pattern is shown in Figure 5.3b.

[1] `https://darkpatterns.uxp2.com/patterns/`, last accessed: 2024-08-07



(a) Dark Pattern (b) Bright Pattern

Figure 5.3: Toying with Emotion Mockups

5.1.4 Summary of Findings

The proposed patterns are bright patterns which prioritize users' long-term goals over short-term desires or business goals. They do not aim to change social media in first line, but add features that aim to change users' behavior in order to work against the problems faced while using social media.

The three design pattern ideas answer Research Question 2 (section 1.2).

5.2 Assessment Results

This section presents the findings from the assessment of the bright patterns as well as the dark patterns. First, the organization of the assessment setting and the participants' demographics are presented, followed by the findings.

5.2.1 Organisation and Demographics

The pattern ideas were assessed within the first two weeks of June 2024 in Bloomington, Indiana (USA). Participants were invited to participate in an available room at Indiana University of Bloomington or online. Three participants took part online via Zoom as they were unable to come to campus. In each study session, the only attendees were one participant and the researcher of this thesis. Participants could fill in the UEEQ on a MacBook or an iPad, as preferred. In the meantime, the researcher was sitting in silence in order not to disturb them but did stay in the same room in case participants had some questions. After this first part, the researcher asked the participants questions to dive deeper into their perceptions and opinions, as previously described in section 3.5.1. Each interview lasted, on average, half an hour, with some being 20 minutes and some up to 40 minutes, depending on how talkative a participant was. It was conducted in an informal and open way.

Participants assessed the following six patterns during the user study:

- Nagging Dark Pattern
- Nagging Bright Pattern
- False Hierarchy Dark Pattern
- False Hierarchy Bright Pattern
- Toying with Emotion Dark Pattern
- Toying with Emotion Bright Pattern

In total, ten participants took part in the study. However, P3 was not familiar with social media and common social media design patterns (in a survey response, they wrote they did not know the term social media feed; in the interview, they said they were not very familiar with social media). And since the evaluation was based solely on mockups without some interactive tools, it was hard for them to understand the patterns they had to evaluate. This resulted in short survey answers and *"don't know"* interview answers. Therefore, this person's data was removed due to a lack of contextual knowledge about social media. This resulted in a final data set of nine participants.

The nine participants grew up in six different countries on three different continents: two in the US, two in China, two in Pakistan, one in Myanmar, one in Russia, and one in Germany. Since only two participants were originally from the US, more than three-quarters participated in their second language. They were in their twenties, and one in their early thirties, ranging from 22 to 32 years. Two-thirds were male, and one-third were female. Five of them considered themselves technically savvy, three a bit, and one did not consider themselves technically savvy.

5.2.2 Findings: Perception of Dark and Bright Patterns

The findings derived from the mixed method study, including the evaluations on the UEEQ and the answers from the interview, are described in the upcoming sections.

Dark and Bright Patterns are Not Attractive

All patterns have a negative mean on the *Attractiveness* scale, except the *Toying with Emotion* bright pattern, which has a neutral mean (0.00), as illustrated in Table 5.1. In the interviews, participants used words like *annoying*, *frustrating*, *pressuring*, *deceptive* among others to describe the patterns.

Pattern	Attractiveness	Pressure	Addiction	Covert	Deception
Nagging DP	-2.24	-2.67	1.00	1.00	-0.71
Nagging BP	-0.94	-0.22	2.00	2.00	1.38
False Hierarchy DP	-0.88	-0.33	1.00	0.11	-1.44
False Hierarchy BP	-0.50	0.22	0.50	0.44	1.38
Toying w Emotion DP	-1.35	-1.67	1.40	0.62	-1.62
Toying w Emotion BP	0.00	0.25	1.17	0.38	1.57

Pattern	Dependability	Efficiency	Perspicuity	Stimulation	Overall
Nagging DP	-0.50	-0.47	1.61	-1.94	-0.62
Nagging BP	-0.50	-0.19	1.61	-1.19	0.19
False Hierarchy DP	0.00	0.88	0.83	-0.60	-0.02
False Hierarchy BP	-0.06	0.50	0.39	0.12	0.25
Toying w Emotion DP	-0.29	0.56	0.94	0.00	-0.15
Toying w Emotion BP	0.12	0.41	0.78	0.47	0.48

Table 5.1: Average Scale Means and Overall Mean per Pattern

Dark and Bright Patterns are Annoying

As illustrated in Table 5.2, every pattern scored a negative *annoying* item score. This means they were considered *annoying* rather than *enjoyable*, which was mirrored in the interview responses, where *annoying* was the most commonly used adjective to describe it.

Nagging was perceived as *annoying* because it was distracting the participants by repeatedly popping up. The bright pattern of it was also described as *annoying* due to its judgemental statement of commanding users to be in the moment. Participants characterized the patterns of *Toying with Emotion* to be *annoying* because of the wording, guilt-tripping the users, and inducing shame. In contrast, the *False Hierarchy* bright pattern was perceived as *annoying*, especially when users want to upload personal content that does not need to be fact-checked, and the dark pattern was described as *annoying* because the users do not want to share their address book with the social media apps.

Thus, while the participants talked about the patterns of *Nagging* and *Toying with Emotion* as being *annoying*, for *False Hierarchy*, participants rather talked about the features themselves being *annoying*.

One person once also said that the manipulating part about the pattern is *annoying*, elaborating on their feelings like,

> P2: *"Annoyed, because the pop up is trying to manipulate me into not deleting my account. It is deceptive and undermines my decision."*

	Attractiveness		**Dependability**		**Efficiency**
	annoying vs enjoyable	**unfriendly vs friendly**	**unpredictable vs predictable**	**obstructive vs supportive**	**inefficient vs efficient**
N DP	-3.00	-1.38	0.71	-1.44	-0.44
N BP	-2.00	0.43	-0.22	-0.78	-0.56
FH DP	-1.25	-0.50	1.25	-1.11	0.88
FH BP	-1.00	0.00	-0.57	0.33	-0.22
TE DP	-1.11	-1.63	1.11	-1.88	0.33
TE BP	-0.67	0.86	-0.38	0.63	0.75

	Efficiency	**Perspicuity**		**Stimulation**	
	cluttered vs organized	**complicated vs easy**	**confusing vs clear**	**boring vs exciting**	**not interesting vs interesting**
N DP	-0.50	1.78	1.44	-2.00	-1.89
N BP	0.29	1.33	1.89	-1.50	-0.88
FH DP	0.89	0.67	1.00	-0.29	-0.88
FH BP	1.22	0.00	0.78	-0.50	0.67
TE DP	0.78	0.89	1.00	-0.14	0.11
TE BP	0.11	0.78	0.78	0.13	0.78

Table 5.2: Average Item Scores per Pattern. N = Nagging, FH = False Hierarchy, TE = Toying with Emotion. DP = Dark Pattern, BP = Bright Pattern

Nagging is the Worst Pattern

Nagging scored the lowest overall mean and lowest mean on the *Attractiveness* scale, as shown in Table 5.1. In the interview, some participants commented that these popups would make users feel like they had an option but that this was actually not the case. They mentioned that the only way to stop this would be to accept it, even though that would not align with what the users want. Otherwise, the app would just *"keep ramming it into your face"*. They complained about the available options, particularly that there is no option to decline it. They consider these options as already manipulative. One person said,

> P9: *"they don't give you a 'No' option, they always give you like, 'no, not now', 'not yet', implying that eventually you're gonna succumb to this. And*

like, while this in itself is kind of manipulative."

Moreover, the *Nagging* dark pattern scored a -3.00 and the *Nagging* bright pattern a -2.00 on the *enjoyable/annoying* scale (Table 5.2), meaning that all participants consider *Nagging* as *annoying*. The reoccuring of the popup made participants annoyed and angry regardless of the intention behind the popup, with one person writing in capital letters that they did not want to be asked again. They think it is manipulative and especially an *"annoying way to tell you what you should do"*. The popups would clutter the screen and interrupt enjoyable moments. The bright pattern popup was, particularly, considered as a moment disruption. One person imagined a situation when they would want to film the moment when Taylor Swift enters a stage, having paid lots of money to see her live at a concert and then failing to take the best-focused picture due to this popup. This would ruin their night. Another person said that this would stop them from using the filming feature as the *"user experience is not smooth enough"*.

Furthermore, this pattern was perceived as patronizing. One participant complained about the dark pattern that it *"undermines the user's choice and intelligence to turn on the notifications by themselves"*. About the bright pattern, two participants said,

P4: *"I decided to film a longer time. And I know when I should stop it"*

P5: *"I think most people when they take a video they are, they're very aware of what they're doing and what they need"*

Toying with Emotion Bright Pattern is the best one

The *Toying with Emotion* bright pattern scored the highest overall mean and a neutral mean on the *Attractiveness* scale, as depicted in Table 5.1. When asking participants to elaborate on the patterns that positively or negatively stood out, most participants talked about this one being the best as it is less intrusive, one person saying *"it looks more like a real post on social media"*. Additionally, it would be thought-provoking and give users a wake-up call to reflect on themselves and think about how much time they had already spent and think about stopping browsing. Participants think it is helpful by describing it as,

P2: *"A friendly suggestion helping me avoid doom scroll the same content"*

P5: *"this kind of pop up may help me to not get addicted to social media"*

One person, on the one hand, felt that the intention was good but at the same time compared it with their mom telling them what to do. Another person who also described it as interesting and thought-provoking complained about it being patronizing, stating that social media should not create some guilt but just let people use it. This person was the only participant perceiving this pattern as manipulative. Others described it as a neutral statement that they had seen before, as not forcing but just a helpful reminder, as giving the user a choice, and as an attempt to manipulate but failing at it due to not being attention-catching enough.

Even though most participants liked it, some were a little bit confused by it. They were not sure what the *"same content as always"* referred to; if it was a hidden picture or the following pictures. It was also not clear if it was clickable. One person thought it was some kind of clickbait where they intentionally put too little information so that the user feels confused resulting in a desire to resolve the confusion and clicking on the picture, saying that they thought that someone was trying to sell them something:

> P9: *"like the way I read it is like, you know, 'your feed is garbage like the real content is here, just click just tap on it, and you'll get the real content.'"*

Bright Patterns are Perceived More Positively than Dark Patterns

All bright patterns have a positive overall mean, while all dark patterns scored a negative overall mean (Table 5.1). Additionally, they have a higher score on the *Attractiveness* scale than their corresponding dark patterns.

Looking at the single items in Table 5.2, bright patterns are *friendlier, less annoying, less obstructive*, and *more interesting* than dark patterns. In the interview, only one person mainly used positive adjectives to describe dark patterns. They used *understandable* and *self-explanatory*. In contrast to that, several participants used positive adjectives to describe bright patterns, like *useful, helpful, cool*, and *beneficial*.

Bright Patterns are Less Deceptive than Dark Patterns

Comparing the corresponding dark and bright pattern pairs on each scale, all bright patterns are better on the *Deception, Pressure, Attractiveness*, and *Stimulation* scales, as shown in Table 5.3. Bright Patterns are at least 2.09 but up to 3.19 *less deceptive*, which coincides with how participants described them in the interviews. While any version of the word *deceptive* was used for each dark pattern at least once, it was only used for the *False Hierarchy* bright pattern.

Moreover, some participants mentioned in the interview that the dark pattern of *False Hierarchy* is *"the evil one"* (P7), implying that they recognized the same technique with bad intentions. One person said,

> P2: *"the same thing that I said before, the other option is not highlighted enough. And in this case, the intention seems nefarious not good."*.

Dark Patterns are More Predictable and Less Complicated

Table 5.2 presents that each dark pattern is *more predictable* and *easier* than its corresponding bright pattern. In the interview, participants were curious about what happens after an interaction with the *False Hierarchy* bright pattern, implying that it is unpredictable. One person said that the explanations were good and clear but that they would be interested in what happens next and what they would need to do. Another person assumed that many users would not read the text between the heading and the button and then would be confused once they were redirected to papers related to the topic.

Pattern Technique	Attractiveness	Pressure	Addiction	Covert	Deception
Nagging	1.30	2.45	1.00	1.00	2.09
False Hierarchy	0.38	0.55	-0.50	0.33	2.82
Toying with Emotion	1.35	1.92	-0.23	-0.24	3.19

Pattern Technique	Dependability	Efficiency	Perspicuity	Stimulation	Overall
Nagging	0.00	0.28	0.00	0.75	0.81
False Hierarchy	-0.06	-0.38	-0.44	0.72	0.27
Toying with Emotion	0.41	-0.15	-0.16	0.47	0.63

Table 5.3: Differences between Dark and Bright Pattern Means per each Scale. Values < 0.00: Dark Pattern scored Better, Values > 0.00: Bright Pattern scored Better, Values $= 0.00$: Dark and Bright Pattern scored Equally.

Room for Improvements of Features

Without specifically asking, individual participants suggested some improvements of the patterns.

Regarding the *Nagging* dark pattern, participants desired that this popup was shown only once in the beginning when they started up the app and then did not ask anymore.

And for the *Nagging* bright pattern, participants asked for other button options and less judgemental text, saying,

> P5: *"I think if you have a button like the 'Not now, don't remind for until my last record' or something like that, it would be kind of better."*

> P7: *"I would, like, change this to be less of like a live-in-the-moment text and more of like 'you can't upload longer than 10 seconds'. So, like, just more factual and less, this feels like there's a big judgment in like, how it's worded."*

The options of the *Toying with Emotion* dark pattern should just be *"Yes"* and *"No"* or the pattern should in general use a politer language,

> P1: *"you could have a politer, polite version of it, I think you could have said, 'Oh, your friends will miss you'. Rather than 'you don't care about your friends', you could have pressurized people by saying, 'Oh, your friends will', rather than 'forget you', you could have said, 'will miss you'".*

Besides the confusion that the *Toying with Emotion* bright pattern created, which needs to be resolved by design adaptions, one participant explicitly suggested a message telling the person how much time they had spent on the app. Another person suggested it should be more attention-capturing with *"a strong color design and audio behind"* in order to be manipulative.

For both *False Hierarchy* patterns, users desired a version with two equally designed

buttons and better button text so that the user knew what would happen without reading the text between the heading and button.

5.2.3 Summary of Findings

Summarizing the aforementioned points, there are two main findings, answering Research Question 3 (section 1.2):

- **Dark and Bright Patterns are Perceived as Nuisance**: Overall, users do not like the patterns shown to them in the mockups, as they are *annoying* and *not attractive*, with *Nagging* being the worst one. Participants spoke about suggestions to improve them.
- **Bright Patterns are Better than Dark Patterns**: Except for dark patterns being *more predictable* and *easier*, bright patterns were overall rated better. Bright patterns scored better in the interview and on the *Deception, Pressure, Attractiveness*, and *Stimulation* scales.

CHAPTER 6

Discussion

This thesis explored how bright patterns can be applied in social media and investigated both dark and bright patterns from an end-user perspective. While earlier studies have explored end-user perspectives on dark patterns [34, 50] as well as the efficiency of bright patterns [37, 78], they have not explicitly addressed end-users' perceptions of dark and bright patterns in the context of social media.

The following sections combine and discuss the research, design, and assessment findings to answer this thesis' research questions. Subsequently, ethical and design dilemmas that occur when conceptualizing bright patterns are discussed. Afterward, I list design guidelines based on the findings. In the end, limitations and potentials for future work of this thesis are presented.

6.1 Research Questions

This section discusses the three research questions of the thesis by examining the findings and their interpretations. However, before delving deeper into the answers to the research questions, I want to highlight the importance of distinguishing between dark pattern type and dark pattern instance due to the versatility of one dark pattern type. One dark pattern type can be implemented differently in order to achieve different goals, and they can be used for good or ill [36]. Accordingly, the perceptions of users are different depending on the pattern, which was shown by the dark pattern evaluation of *Hindering Account Deletion*. It scored neutrally in the survey of this thesis, while it was negatively experienced in other research [34]. I assume that *Hindering Account Deletion* might have been evaluated differently if, for example, users were required to take several steps to delete their accounts. Consequently, this suggests that dark pattern types can not be generalized. The subsequently discussed findings are understood on an instance level rather than on a type level.

6.1.1 What are known problematic design patterns that manipulate users on social media, and how do users perceive them?

Social Media Manipulates in 15 Different Ways: In this thesis, I identified 15 different known low-level dark patterns, categorized into six high-level dark patterns, commonly utilized by social media companies. In the process of creating my taxonomy, I realized that dark patterns can not generally be categorized into different social media company goals due to their versatility. One single dark pattern technique can be applied to achieve different goals. For instance, *Toying with Emotion* can be utilized to manipulate users into sharing more personal information as well as to manipulate them to stay on the social media platform. However, by further investigating users' perception of these 15 dark patterns, I found that low-level dark patterns within the high-level dark patterns *Interface Interference* and *Social Engineering* are perceived similarly with an average correlation of up to 0.77. This verifies that dark patterns can be categorized into similar techniques. Additionally, the identified six high-level dark patterns align with the six high-level dark patterns identified by Gray et al. [35].

Deceptive Dark Patterns are Ugly: The user perceptions investigation results showed that there is a strong correlation between the *Deceptive* and the *Attractiveness* scale, which indicates that *deceptive* dark patterns are *not attractive*. This aligns with another study's finding, where the incorporation of dark patterns led to a significant difference on the *Attractiveness* scale [17]. In yet another study, they found a significant association between the characteristics *asymmetric* and *restrictive* with the *malice score* [54]. The survey results additionally indicate that dark patterns are *not attractive, pressuring, deceptive, annoying*, and *unfriendly*. Earlier studies found that malicious techniques are *frustrating* [21].

Forced Access is The Worst Snake In The Grass: While participants in other research considered *Forced Action* more acceptable than other dark patterns [50], my survey results showed that the *Forced Access* dark pattern was perceived worst, scoring the lowest overall mean and perceived as the *most covert* one.

Some Blessings In Disguise: Some dark patterns were not exclusively evaluated negatively but were also perceived positively in some aspects. For example, *Gamification* was rated as *pressuring* but also *easy*. This partly aligns with the statement of Gray et al. that stated that some dark pattern instances, like *Forced Action* and *Nagging*, perform well in usability testing [36], and amplifies the idea of exploring how dark patterns can be used for good.

6.1.2 How could known dark pattern mechanics be repurposed as countermeasures against users' perceived social media problems?

This question implies knowing what users perceive as problematic of social media as well as what they would like to change. By investigating users' perceived social media problems and ideas for changes, eight main problems arose, with three being focused on, namely Social media is time-consuming and time-wasting, Social media spreads fake

news, and Social media does not provide meaningful content.
I decided to go with the three low-level dark patterns that were evaluated as least attractive in the survey (*Toying with Emotion*, *False Hierarchy*, and *Nagging*), making the challenge of finding a positive narrative even more difficult. Three ideas for bright patterns were elaborated, as described in the following. All the pattern ideas introduce friction into the user experience as they are proven to have positive effects, for instance, supporting health behavior change [23], or reducing the time spent on Facebook while enjoying it more [65].

Nagging Users into Reducing Time: The first bright pattern idea is to use *Nagging* to reduce users' spent time on social media. In the survey, *Nagging* was perceived particularly *pressuring, obstructive*, and *annoying*. The newly created bright pattern used the same technique in order to push users into stopping using social media. It does not force the user to leave the app by restricting them in any way; rather, it is thought-provoking and thus motivates users to leave. In practice, this is a popup showing up while filming a sequence, encouraging them to stop filming and rather enjoy the present moment. This bright pattern idea also aligns with the idea of repeating the call to action with the use of popups, sidebars, or top bars so that it is always present in the users' sight in order to persuade users, as mentioned by a designer in another study [75].

Falsely Hierarching the Option to Fact-Check: The second bright pattern idea is to use *False Hierarchy* to decelerate the spread of fake news. In the survey, *False Hierarchy* had been evaluated as *unfriendly, pressuring, deceptive*, and *annoying*. The same technique was used to encourage social media users to fact-check content before they upload or share it. This idea is not based on deleting any social media posts that are already online but rather on encouraging users not to post stuff that was not fact-checked in the first place.

Toying with Users' Emotions to Change the Content Algorithm: The third bright pattern idea is to use *Toying with Emotion* in order to change the input of the content customization algorithm. *Toying with Emotion* had been evaluated as *annoying, unfriendly, deceptive*, and *pressuring*. The new bright pattern utilizes the same technique in order to blame users for looking at the same content as always, thereby encouraging them to look for different content, which subsequently leads to a different input for the content algorithm. Different content helps in presenting meaningful content as well as lowering the distortion of reality created by seeing the same content all the time. This pattern would additionally decrease the time spent as it serves as a reminder or as an interruption while scrolling through the social media feed. This pattern does not change social media's content algorithms but rather inspires users to explore different content by motivating them to rethink what content they are looking at. It tries to make them aware of the fact that they fell into a rabbit hole and, therefore, always get displayed the same kind of content.

Three Reasons for Bright Patterns: There are three main reasons for incorporating bright patterns. Firstly, the incorporation of bright patterns is considered a useful help to develop ethical good software [37]. The utilization of dark patterns may be good for

short-term gains but not for long-term goals. On the one hand, users will leave the websites as soon as they find a better website for the same product or service [78]. Some participants in a different study also reported feeling valued for their money, and not valued as a person [34]. On the other hand, employees will leave the company if they do not align with the ethics guidelines of the company. Thus, the company not only loses profit but also knowledge [78]. I assume that the incorporation of bright patterns attracts users and employees.
Secondly, Monge Roffarello and De Russis argue that disclaiming the utilization of dark patterns may lead to less user engagement and, therefore, less revenue in the short term but could increase user loyalty in the long term [57]. Conversely, I assume that using bright patterns could increase user loyalty.
Thirdly, Susser and Grimaldi emphasize concerns regarding automated influence due to the current increase of artificial intelligence and machine learning algorithms [74]. It is important to build more and more bright patterns that fill the artificial intelligence data pool with influential patterns for the benefit of users.

6.1.3 How do users perceive the newly designed bright patterns compared to the corresponding dark patterns?

All Patterns are Nuisance: I found that dark as well as bright patterns are perceived as nuisance. They were described as *annoying* and *not attractive*.

Bright Patterns are Better than Dark Patterns: This thesis demonstrated that participants favored bright patterns over their corresponding dark patterns by achieving a better overall mean. This finding aligns with a previous study where participants preferred the prototype incorporating bright patterns over the one with dark patterns [78]. In this thesis, bright patterns were perceived as *friendlier, less annoying, less obstructive*, and *more interesting* compared to their dark patterns. While literature lacks in comparing this result, Graßl et al. did not find significant differences between dark and bright patterns regarding perceived control over personal data [37].

6.2 Ethical Dilemmas

This thesis raised two ethical dilemmas. First, it is debatable whether dark patterns should be weaponized for the user's good. Second, it is questionable if bright patterns invade users' autonomy and whether such an approach is acceptable.

6.2.1 Weaponization of Dark Patterns

Throughout the progress of my thesis, I encountered criticisms regarding the idea of weaponizing dark patterns for positive outcomes. I want to clarify that my interest in researching in this area does not imply support for dark patterns. Instead, my goal was to explore a positive application of them, guided by the belief that every coin has two sides. In the following, I will explain why this approach is a legitimate attempt.

According to Schäfer et al. [72], which aligns with participants' suggestions in another study by Maier and Harr [50], there are four ways to deal with dark patterns, namely raising the users' awareness, putting pressure on companies by publicly shaming them for their practices, inventing laws forbidding the utilization of dark patterns, and developing technical dark pattern detectors and countermeasures. In all of these four categories, attempts have been made by different stakeholders but they remain challenging, as listed in the following.

Being aware of dark patterns does not influence the impact these techniques have on users: In a study, users thought they had never been tricked by websites. At the same time, they were not able to identify all presented types of dark patterns and became victims to at least one of them [49]. Bösch et al. argue that while raising users' awareness could lead to them becoming self-motivated to configure their accounts properly, this would be hard to achieve [16]. Simultaneously, Bongard-Blanchy et al. and Maier and Harr argue that raising users' awareness about online manipulative attempts does, on average, not impact the influence these designs have on users [10, 50].

It is difficult to identify the person in charge of the creation of these manipulative designs: Most users think designers, other stakeholders, and developers are to blame for creating manipulative applications [34]. Therefore, people share dark pattern examples on public online platforms in order to shame designers and companies and put pressure on them [18, 31]. But it is difficult to prove designers' and companies' intention behind the creation of dark patterns [41]. A second rationale for not blaming designers is because it is questionable to which degree designers are responsible for the design they create [36]. They are limited to the company's agenda, lack in having a fair system in which they can share their ethical standards without fearing negative consequences, and are not moral philosophers despite having an intuition [75]. Moreover, designers mentioned in a workshop that dark patterns are design choices driven by market forces and, thus, are not under their control[4]. Furthermore, it is sometimes not even the website owners that are responsible but third-party entities that create and present dark patterns on behalf of the shopping websites [52].

There are various challenges in legislatively regulating dark patterns: According to researchers, there is a need for governmental interventions to counteract dark patterns [10], particularly mild dark patterns [46]. However, there are challenges in doing so. Firstly, prohibiting particular interaction designs might lead to designers adapting their designs to align with the law by thwarting prohibitions [41]. Secondly, some dark pattern strategies like *Toying with Emotion* are hard to restrict due to companies' freedom of speech [46]. Thirdly, it is difficult to evaluate designs that are at the border between dark and gray [41]. Also, legal inventions sometimes produce outcomes that are not intended, as seen by cookies. Cookies were invented to discharge legal liabilities, stating that businesses need users to accept cookie tracking before being allowed to track them. Originally, these cookie consent requests are designed to give users greater control over their own data [37] and to inform them [24]. However, in reality, they include dark patterns leading to users making quick decisions [37].

Automatic detectors and countermeasures do not work sufficiently and elicit other challenges: Even though literature presents some promising attempts to create automatic dark pattern detectors, there are still some challenges that seem irresolvable. The different approaches presented in section 2.1.1 have various limitations: some are limited to some kinds of dark patterns, others are not capable of detecting dark patterns taking advantage of style, color, and other non-textual features [52], some are restricted to detect ten different dark pattern instances [42, 51], and others depend on the performance of anti-virus scanners [45]. Moreover, there are the following challenges that make the automatic recognition of dark patterns difficult. Firstly, websites are often not HTML markup and design compliant, creating a need for the identification of different design alternatives and interacting with product options [52]. Secondly, there are different kinds of dark patterns with varying implementation methods [45, 51] and some of them are simply not automatically detectable since there are too many different variations of implementation [25]. Thirdly, these devious design patterns can change over time. Thus, the detection approach should be extensible so that new devious groups can be easily covered [45]. Given this, automatically detecting dark patterns is a challenging task and to the best of my knowledge, there does not exist a promising solution to automatically detect dark patterns.
A similar situation occurs with other countermeasures. Even though literature presents promising countermeasures [65], there are some irresolvable challenges. Companies are continually developing more and more aggressive techniques to bypass the defense mechanisms applied by users [21] and counteracted by technologies [72]. Users should not exclusively trust technology to protect themselves [72].

As a result, these four different approaches, especially the latter two, are always at least one step behind. Given these challenges in regulating dark patterns, and considering that these manipulative techniques are already in use, why not repurpose them for the benefit of users?

6.2.2 Autonomy of users and acceptance of manipulation

Autonomy needs to be ensured by participatory design: Even though another study showed that a prototype with bright patterns was perceived as providing higher freedom of choice and higher transparency than a prototype with dark patterns [78], the user autonomy is still questionable. This is also something that fair patterns want to address [63]. Also, the controversial thing about bright patterns lies in their definition of prioritizing users' goals. But what are users' goals? Is the user's autonomy maintained if the designer just assumes what the user's goals are? I do not think so. User goals should be defined by the users themselves, either directly in the system or by user research. In my case, the survey helped to understand the user's goals, laying the foundation for creating solutions and developing bright patterns.

Perceived manipulation depends on how participants define this term: When it comes to the acceptance of manipulation, it is important to understand how participants define

manipulation. As detected in the interview, manipulation is understood differently among participants. Here are some different statements:

> P2: *"Manipulation means you do not, you are undermining, you are not respecting the intelligence of the user, respected, they will make their own choices. They know how to make choices. If you are manipulating them, it seems like you are trying to tell them 'hey you don't know what's right for you.' Even if they don't know what's the right for them. It's their choice. They'll do it by themselves"*

> P4: *"if you really restrict me to doing something, I will consider it as manipulative"*

> P5: *"for me manipulating is someone or some in these cases, social media app, they use their power or they use what they have to makes you do what they want you to do. Yeah. But maybe in like a strong way, or in a not friendly way"*

> P8: *"it's when someone tries to convince you that like the best for you to do it is that way or to do something in general and but they actually, It's not about"*

While one participant defines manipulation as convincing someone to do something, another participant described it as restricting their actions. This person also thought that *Nagging* was the most manipulative. At the same time, they said that *False Hierarchy* is only mild manipulation since even though the design encourages the user to click the highlighted button, they still have a choice. This participant considered this mild manipulation acceptable. Another participant perceives manipulation as undermining and not respecting one's intelligence since the person manipulating is making choices for the other, thinking that they know better what is right for them, even though they don't. In the data from these nine participants, Chinese people perceived something as more manipulative the more forcing it was, while other people perceived something as more manipulative the more deceptive it was. This tendency might differ with a larger sample size.

Manipulation is acceptable if it is for the greater good: Three participants said that manipulation is okay when it is for the greater good, with one saying:

> P10: *"I think as long as it was in the best interest for everyone, and was equally fair and protective over everyone, then that's okay to kind of manipulate people into following that".*

One person said that politeness could also be considered manipulation in a broader sense. Another person said that highlighting some often-used buttons in a design could also be considered manipulation. Both these two would be okay by the participants. One opinion was that it is okay to manipulate kids and that manipulation, in general, would not be too bad, but it needs to be balanced depending on the language and the purpose. However, manipulating social media users would not be acceptable. Other answers were

according to shown patterns, with one person saying that a mild *False Hierarchy* would be okay, another one saying that a *False Hierarchy* with clear button text would be okay, and another one saying that *Nagging* with only one popup would be okay. Two participants said that it's never okay to manipulate people. Interestingly, no participant said that manipulation would be okay if they could customize it, implying that none of them believed this was an option.

The question remains: Is it important to not invade one's autonomy when it is effective?

6.3 Design Dilemmas

When designing bright patterns alias *"persuasive user interface design functionality that prioritizes users' well-being and goals over their desires and business' objectives"* [67] some design dilemmas arise, as elaborated in the following.

6.3.1 Definition of Bright Patterns

The definition of bright patterns says that these techniques prioritize user's goals over desires, and Ozkaramanli et al. introduced six different design strategies to achieve that. Three strategies demotivate immediate desires, and three strategies motivate long-term goals [62]. Analyzing my proposed bright patterns according to these strategies, I uncovered that the *False Hierarchy* bright pattern did not use any of these strategies. On the contrary, it enables desires by making the *"Share without fact-checking"* an easier and faster procedure. It might have been good to make gains tangible, for example by saying that fact-checking will reduce the spread of fake news. The other two patterns incorporate proposed strategies by Ozkaramanli et al. *Nagging* adds new sources of displeasure to demotivate temptation, whereby the popup serves as displeasure. *Toying with Emotion* makes potential losses of temptations tangible by saying that the content is garbage. I assume that bright patterns that motivate long-term goals would be perceived better.

Coming back to the definition of bright patterns, focusing on the part of prioritizing goals over desires already raises the question whether such patterns can be perceived positively in first impression assessments. I assume that they very likely result in negative evaluations when they violate desires.

Moreover, the definition says that user goals are prioritized over company goals. In the user study, one participant questioned the practicality of the created bright patterns, especially of the *False Hierarchy* one. Also Bongard-Blanchy et al. request governmental interventions to counteract dark patterns, as it is an unrealistic belief that companies would implement countermeasures out of their interest [10]. Similarly, it is probably a utopian thought to believe that companies would implement bright patterns.

The questions are: Do realistic bright patterns exist? Can bright patterns be perceived positively in first impression assessments or are long-term studies needed?

6.3.2 Attractiveness of behavior change

As concluded from the assessment of the dark and bright patterns findings, none of the patterns were *attractive*, but all of them were *annoying*. *Nagging* was perceived worst among the three patterns, scoring the lowest overall mean and the lowest mean on the *Attractiveness* scale. Participants perceived both the bright as well as the dark pattern as particularly *annoying*, which aligns with findings of other studies [50]. However, Neyman found in their study that *Nagging* users every five minutes with a prompt if they want to continue scrolling leads to a reduction of 50% in time spent [60].

Regardless of the possibility of bright patterns achieving high scores in first impression assessments, as elaborated in the prior dilemma, the question remains whether they do need to score well in user experience evaluations. For example, frictions also do not win when it comes to ease of interaction. However, when they are introduced with intention, they can have positive effects on behavior change. Cox et al. argue that designers need to weigh trade-offs between creating efficient everyday interfaces and making users more mindful of their behavior [23].

On the contrary, Fogg et al. say that the more attractive an interface or hardware is to the target group, the more power it has to persuade. But if the target group is not accordingly addressed, the users may get irritated or angry. Designers need to know the target group for the product in detail to persuade them [33]. Similarly, Ozkaramanli et al. emphasize the importance of detailed exploration and careful analysis of the motivations underlying specific behaviors in order to design solutions to self-control dilemmas [62].

The questions remain: To what extent is negative attractiveness acceptable in order to achieve a desired behavior change? How much burden on users is justifiable? Does it matter how users perceive these patterns, if they are effective?

6.3.3 Finding the perfect dose of manipulation

There are some rules of thumb that enlarge the power of persuasion, namely, the more human-like, the more attractive, the more similar to the user, and the more praise-giving the software is, the greater power it has to persuade the user. Furthermore, users tend to reciprocate in response to software. This implies that if users perceive a benefit, they are open to returning the favor. Designers can fully utilize this by, for instance, displaying a message like *"You have enjoyed playing this game ten times. Why not pay back the favor and register?"*. Also, the social role that software mimics influences the power of persuasion. Especially authority roles have great power to change a user's behavior, and so do roles of friend, entertainer, and opponent[33]. Luguri and Strahilevitz stated that too aggressive dark patterns could lead to customer backlash [46]. This was proven by the user study, which revealed that patterns often trigger negative emotions in users, which sometimes result in actions not aimed by the pattern creator. For instance, two statements were,

> P1: *"So it's actually gonna annoy me. And then I'll, I'll not make an account on that platform. Just because of this notification."*
>
> P2: *"Super annoyed, I'd just start pressing on the YES, YES thing. As soon as I saw that, because I don't even, I read the first like, line that is in bold. And I'm already annoyed, and I just can't see this. I don't want to read the next thing, because it's super annoying."*

Thus, using such patterns can lead to the user taking an undesired action. Such scenarios were mentioned more often when discussing dark patterns than when discussing their counterparts.

As already mentioned, ethics is subjective. Thus, a designer can think that a bright pattern is ethically good, but it is only if the consumer also perceives the bright pattern to be ethically good. The line between ethical and unethical is very thin [78]. Conti and Sobiesk wrote about a tolerance threshold that designers should not overscore. Up to this point, users tolerate malicious design, but upon this threshold, users will look for alternative ways to complete their desired task. Thus, designers aim to produce manipulative patterns that are aggressive enough to breach the cognitive barriers but, at the same time, not too aggressive to scare users off. Based on their study conducted in 2010, the threshold varies depending on the content, being lower for web searching tasks or seeking vendor technical support and higher for gaming, shopping, sporting, and pornographic websites [21].

The design also needs to find the perfect dose of the provided information. As in a different study presented, users want to have information about the presence of dark patterns while also not liking visually cluttered interfaces due to their impact on their productivity [72].

So the questions are: Until what point is manipulation effective and until what point is it okay?

6.3.4 Designing for social media

As seen in the interviews, there is a particular bias against social media companies that lead to different actions. Participants believe social media is untrustable, especially in regards to collecting data, as in the attempt in the *False Hierarchy* dark pattern. The following two statements illustrate that:

> P6: *"I mean, even though I'm not sure if it is, like, very accurate, or, I mean, if they are still collecting your data, even though you said no"*
>
> P10: *"But sometimes I feel like that's not entirely true that because if you do sign up with your email or your number, and you still choose not to connect with your contacts, I feel like that it still does anyways, kind of like without your consent."*

And the general social media bias is demonstrated in the following:

> P6: *"I mean, in general, like, everything on social media, whatever the feature, could be considered as manipulative"*
>
> P7: *"Like I never actually want to click on the thing that they're clearly trying to get me to click on. And maybe that's because I feel like at odds with like social media companies."*

This suggests that some users do not trust social media in first place. Can bright patterns be designed in the context of social media?

6.3.5 Innovation of Bright Patterns versus Usability of Dark Patterns

In a study, one designer followed the rules of usability and said that sometimes similarity to known things is preferred over innovation since you could create confusion [75]. Simultaneously, Maier and Harr said that participants' responses to dark patterns depend on previous personal experiences [50]. These two statements could explain why dark patterns were perceived as more predictable and less complicated than their counterparts in this study. Users are already adjusted to some dark patterns and they are influenced by acquired expectations from earlier interactions. Thus, they do not have to think a lot and just act automatically. This is illustrated by a participant who did not see the text beneath the button in both *False Hierarchy* patterns. While they complained about the bright pattern lacking in the possibility of skipping the fact-checking option, they did not complain about the dark pattern because they just assumed that there would be an option of signing up without uploading their contacts. This example illustrates that this person did not read the text, and just by knowing this kind of dark pattern, they already knew what it was about. On the other hand, they did not know the bright pattern and thus were not aware that the option of skipping existed. Some participants mentioned other forms of automatic behavior, such as just scrolling without thinking, but also automatic behavior that works against dark patterns, like just declining popup prompts or clicking the thing in fine print. Compared to dark patterns, the proposed bright patterns would be rather new, and thus, users do not yet have such mechanisms to automatically take action. This is also seen by the fact that each bright pattern somehow troubled confusion. Some participants were confused about whether the video was still recorded while the *Nagging* popup was displayed, they were confused about the next steps after clicking the button in the *False Hierarchy* pattern, and they were confused about the design of the *Toying with Emotion* bright pattern.

And Truong and Dalbard suggest that people have already accepted the existence of dark patterns and are using automatic mechanisms to reach their goal as fast as possible [78]. Diving deeper into this, let us take on a thought experiment: Upon opening a website in the browser, the user sees a popup asking for cookie consent. Because they are in a rush, they do not read but immediately recognize one obtrusive and one unobtrusive button. As they have already come across such dark patterns, they know of them and immediately click on the unobtrusive button. Now let us analyze this: If this website implements a dark pattern, it was the best option the user took, but if the website implements a bright

pattern, then the user made the wrong decision because they clicked on the button to accept the cookies.

The questions are: Can bright patterns be perceived positively when they are new? Would they rather confuse users, or even lead to undesired outcomes?

6.4 Design Guidelines

The ideas presented in the following are insights gained from this research. I share them as potential foundations for creating new bright patterns, but they should not be seen as a comprehensive list of design guidelines.

6.4.1 Bright Patterns Should Not Be Patronizing

In the interview, some participants criticized patterns in a way that implies that they believe they are patronizing. For instance, these are two statements by two participants:

> P1: *"If I am sharing a post, I do not want FB to tell me what to do"*
>
> P8: *"I felt like the intentions were good, but I would be like, it sounds like something that my mom would say."*

Both these statements sound like participants did not like what they were talking about, P8 compared the *Toying with Emotion* bright pattern with their mom. Fogg described technology that assumes a specific role, such as the role of a teacher, a friend, or an entertainer, is more persuasive. The author also mentions that the best-suiting role needs to be chosen in order to not be counterproductive [33]. Albert verifies this as they found out that participants dislike the inappropriate use of familiar tones, for example acting like a friend or parent, and prefer a neutral tone [6].

As the study showed, the *Toying with Emotion* bright pattern was perceived as the best pattern. Participants also called this pattern as *least intrusive*. As a designer of bright patterns, one needs to be aware of the fact that patterns are considered a nuisance, as also demonstrated in this study. This suggests, that bright patterns should not be patronizing, users still value their autonomy, as also mentioned in the open-ended questions in the survey.

6.4.2 Bright Patterns Should Be Thought-Provoking

The incorporation of friction in the newly created bright patterns may have led participants to perceive and describe them as thought-provoking, given that research has shown that friction or obstruction leads to self-reflection [23] and is effective [46].

Besides perceiving the patterns as thought-provoking, one participant also changed their opinion in one single sentence without me intervening just by rethinking about it:

> P1: *"I don't think it's manipulative. It is kind of, actually, it is."*

6.4.3 More Research on Target Group and Goal is Needed

Brooke criticizes the generalization of design features as one feature can be very useful in a specific environment but might not be as useful in a different environment [15]. Similarly, the perception and effectiveness of one pattern depends on the person, one participant said,

> P10: *"It just depends on the person. I think that me personally, I don't take things literally. So I will be like, I'll click the Yes, I don't care about my friends. Obviously, that's not true. But I just want to delete the account. But somebody that maybe is a little younger than me that cares about a lot of things like is in a more stage in their life where they need to feel accepted by other people they might be, stop and think about this, like somebody that's in like middle school or high school, where they're like, Oh, I just want to be accepted by all my friends. They might be like, Well, I don't want my friends to forget me. So they might be manipulated into being like, Okay, I'm gonna go back and, you know, not delete the account."*

The *Toying with Emotion* dark pattern was mostly perceived negatively in the interview, while one participant viewed it positively, interpreting it as a friendly reminder. Different opinions to the same pattern are:

> P5: *"And they say your friends will forget you maybe gives me another thing. Oh, yes, I have many friends here in the social media app, should I, should I delete my account now, but I think most of the time, this kind of pop up will, will appear in the last step when you delete your account. And if you delete, yes, I don't care about my friend, that means you want to delete your account, your account will be like deleted in like in seven days, or something like that. I think it's, it's kind of interesting and it's a kind of friendly suggestion"*

> P8: *"it felt like you don't really have like, you don't really have a choice but like, either you hate your friends or you stay on this app"*

The perspective of it being a friendly reminder aligns with other researchers' who stated that such a step could be seen as a security measure to prevent users from accidentally deleting their account due to a mis-click, despite the emotions it might trigger [27]. That said, patterns need to be adjusted to the goal and target group, which needs to be considered when designing bright patterns.

6.4.4 More Research on Features is Needed

Disregarding the manipulative aspects of the patterns, participants also complained about the features themselves. In particular, the *False Hierarchy* bright pattern led to contradicting opinions. Even though some participants liked the intentions behind it, they critiqued the implementation of it. They complained about the fact that the next steps would be unclear as well as the fact that it would put work on users, while other stakeholders should be responsible for the fact-checking. Users do not want to read

through papers due to lack of time or other reasons. Additionally, putting this work on users would not result in the intended results since users have their own biases, leading to biased fact-checking. One person thought of a situation in which a person finds out that the content they are about to post is not true and then deliberately posts it in order to provoke other social media users. Participants do not want to have this kind of censorship, especially when they only want to post something fun that does not need to be fact-checked. One person believes that this feature would lead to distraction and that people would probably post less, which was confirmed by another person who said, they would post less with this feature. This suggests that the ideas of the features themselves and how they are implemented need to be modified.

6.5 Limitations

By prompting survey participants to evaluate dark patterns in the survey, I did not include any feature without any dark pattern incorporated, which could have served as a control value. I thought this was not necessary since I did not intend to check if users could recognize them but rather wanted to know how they perceived them. However, participants might have realized my opinion about these features, which probably resulted in bias. Additionally, the fact that this dark pattern evaluation was at the end of a larger survey could have probably cognitively overloaded and overwhelmed the users.

With three quarter of the survey participants being from Austria, and another 5% being from Switzerland or Germany, 80% of the participants are from the German-speaking area. Thus, there might be a Western bias in the survey data. Additionally, the used language needs to be reflected on. While the survey was created in both English and German, most participants were able to participate in their mother language. However, the user study was only conducted in English, and seven out of nine participants did not have English as their first language, which possibly made it difficult for them to articulate their thoughts and opinions.

Throughout this thesis, I prompted participants to evaluate dark pattern instances on the basis of mockups. This restricted them to a particular amount of contextual information. Only observing mockups is a different experience than actually interacting with the social media platform. In some instances, the participants would have needed more contextual information, for example, the one participant who was removed from the data set due to lack of knowledge. While I tried to create a more natural experience by creating some videos instead of images, I am uncertain whether I chose the best option for each pattern. For example, the *Expectation Result Mismatch* mockup might have also been a video to help participants understand the flow more easily.

6.6 Contributions

This thesis contributes to the growing body of research regarding dark patterns. It identifies and categorizes dark patterns commonly applied by social media platforms,

resulting in a new taxonomy, and provides an understanding of how users perceive these manipulative patterns. It additionally presents a newly created method to evaluate dark patterns, more elaborated in a paper [1] for a workshop on CHI Conference on Human Factors in Computing Systems (CHI 2024)[2]. Moreover, it investigates the relatively under-explored area of bright patterns. Three newly created bright patterns were created and then assessed to understand users perception and acceptance of them. Moreover, it discusses ethical as well as design dilemmas that arise when designing bright patterns in social media.

6.7 Future Work

This thesis revealed that both dark and bright patterns are perceived as nuisance, which suggests that further investigation is needed to adapt the patterns. In particular, the pattern could be modified to enhance effectiveness while being perceived as less negative. In terms of the *Toying with Emotion* bright pattern, it would be important to know in which situations which degree of negative emotion would be appropriate. The findings of this investigation would also be important for the other patterns, especially since patterns sometimes trigger emotions leading to the user taking the unintended option. Another approach would be to investigate how the characteristics of patronizing influence users' perception of the patterns. Would a bright pattern be perceived as less patronizing if participants were able to turn off the pattern? While this thesis focused on how dark patterns can be repurposed, an alternative approach would be to investigate how the problem can be addressed most effectively by bright patterns. Further research is needed to adapt these patterns to better achieve their goals. To do so, researchers could build upon the improvement ideas suggested by some participants.

In contrast, this thesis also demonstrated that users' perceptions of bright patterns are more positive than their perception of corresponding dark patterns, which underlines the potential of bright patterns and suggests further investigation into this topic. The results laid a foundation for future work to investigate how bright patterns can be applied. The next suggested steps are building prototypes and evaluating them by not only investigating users' perceptions of bright patterns but also the efficiency of these patterns. While A/B tests could measure user actions, long-term studies are needed to investigate the effect of bright patterns on user behavior change.

[1] https://ceur-ws.org/Vol-3720/paper11.pdf, last accessed: 2024-08-03
[2] https://ceur-ws.org/Vol-3720/, last accessed: 2024-08-03

CHAPTER 7

Conclusion

As part of this thesis, a new taxonomy of 15 low-level dark patterns grouped into six high-level dark patterns found in social media was developed. By prompting users to evaluate these dark patterns on the UEEQ, this thesis uncovered a correlation between the *deceptive* and *attractive* scores. The more *deceptive* a pattern, the *less attractive* it is. Participants generally do not like dark patterns, evaluating them as particularly *pressuring, deceptive, annoying*, and *unfriendly*. However, while some dark patterns scored negatively overall, they positively scored in some items on the UEEQ, which strengthened the motivation to explore their applicability in repurposing them for the benefit of users. By further exploring three different dark patterns and assessing their applicability in inverting them, three new bright patterns were conceptualized and prototyped. The bright patterns aim to reduce the time spent (*Nagging*), decelerate the spread of fake news (*False Hierarchy*), and present more meaningful and diverse content (*Toying with Emotion*) on social media. In the assessment of these newly created bright pattern and dark pattern prototypes, bright patterns were perceived more positively than their dark pattern counterparts, while all were perceived negatively overall. This thesis indicates that dark patterns can be repurposed for the user's good and also score better in user perceptions than their corresponding dark patterns but are still perceived as nuisance. The creation of bright patterns raises some ethical concerns and design dilemmas resulting from the definition of bright patterns, user perceptions as well as the context of social media. Further studies investigating the efficiency of such bright patterns are suggested, as these findings rely on end-user perceptions solely.

Appendices

A Taxonomy

Authors	Reference	Used or Not Used with Rationale
Brignull (first version 2010)	Website `https://www.deceptive.design/types`, last accessed: 2024-08-07	**Not used** since it is too general, not focusing on social media
Conti and Sobiesk (2010)	Malicious interface design: exploiting the user [21]	**Not used** because it is too general, not focusing on social media
Zagal et al. (2013)	Dark patterns in the design of games [82]	**Not used** since it focuses on game design, not social media
Greenberg et al. (2014)	Dark patterns in proxemic interactions: a critical perspective [38]	**Not used** because they present dark patterns in proxemic interactions and not social media
Bösch et al. (2016)	Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns [16]	**Not used** because they do not specifically use patterns in social media
Neyman (2017)	A Survey of Addictive Software Design [60]	**Not used** since they present design strategies and the author says that the list is not exhaustive
Gray et al. (2018)	The Dark (Patterns) Side of UX Design [36]	**Not used** because it's too general, not focusing on social media. However, it serves as a framework, meaning my taxonomy is mapped to theirs
Montag et al. (2019)	Addictive Features of Social Media/Messenger Platforms and Freemium Games against the Background of Psychological and Economic Theories [58]	**Not used** since they present mechanisms not types
Mathur et al. (2019)	Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites [52]	**Not used** because it is e-Commerce specific and not social media
Monge Roffarello and De Russis (2022)	Towards Understanding the Dark Patterns That Steal Our Attention [57]	**Used**
EDPB (2023)	Dark patterns in social media platform interfaces: How to recognise and avoid them [27]	**Used**
Mildner et al. (2023)	About Engaging and Governing Strategies: A Thematic Analysis of Dark Paterns in Social Networking Services [56]	**Used**

Table 7.1: Popular Dark Pattern Taxonomies and Rationale of Using or Not Using Them

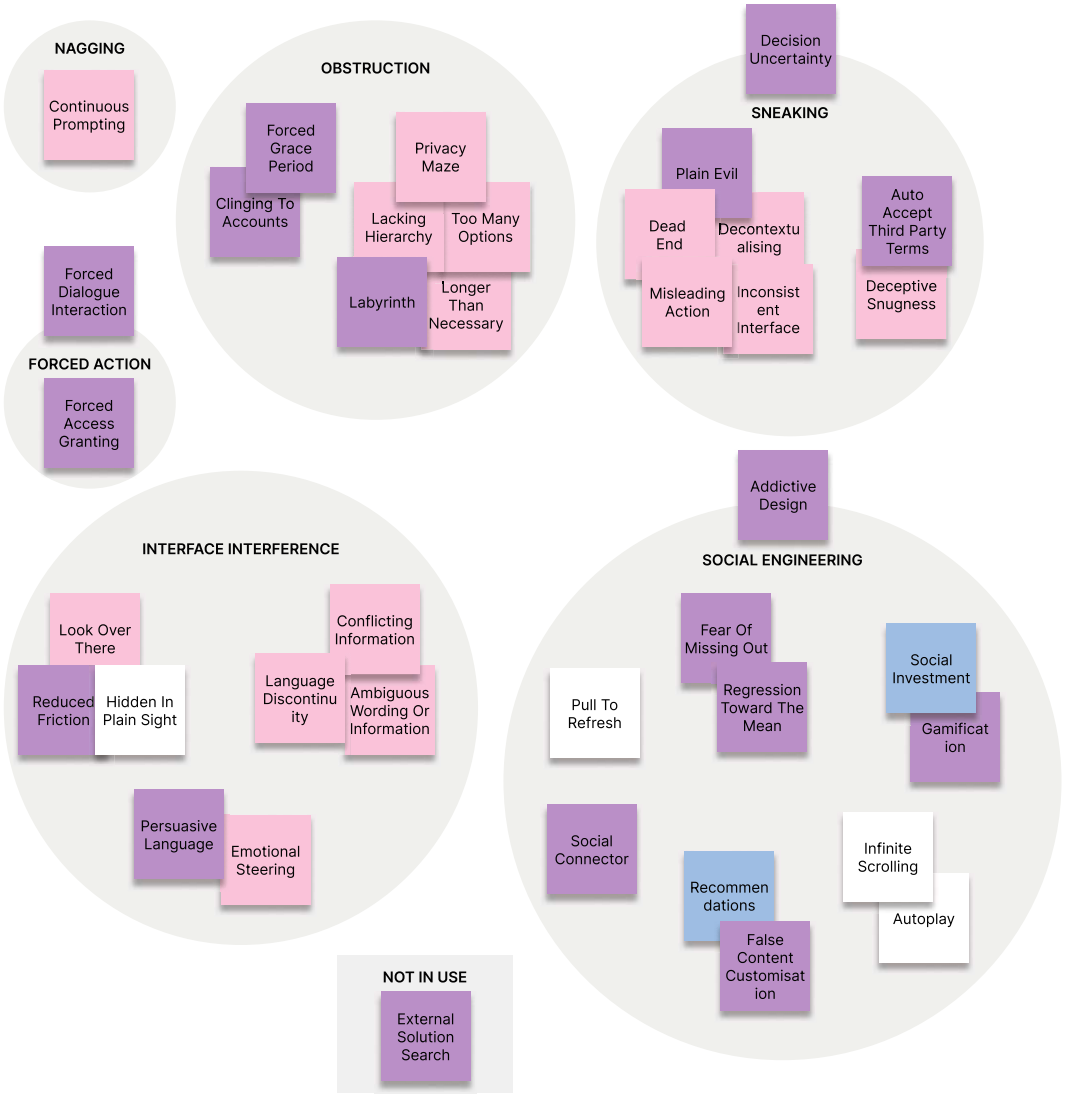


Figure 7.1: Post-its of Finished Taxonomy

Author	Amount	High-Level	Low-Level
Monge Roffarello and De Russis (2022) [57]	1 | 5	Attention capture	Recommendations Autoplay Pull To Refresh Infinite Scrolling Social Investment
European Data Protection Board (2023) [27]	6 | 16	Overloading	Continuous Prompting Privacy Maze Too Many Options
		Skipping	Deceptive Snugness Look Over There
		Stirring	Emotional Steering Hidden In Plain Sight
		Obstruction	Dead End Longer Than Necessary Misleading Action
		Fickle	Lacking Hierarchy Decontextualising Inconsistent Interface Language Discontinuity
		Left in the dark	Conflicting information Ambiguous Wording Or Information
Mildner et al. (2023) [56]	5 | 22	Engaging Strategy - Interactive Hook	Addictive Design Autoplay Content Fear Of Missing Out Gamification Infinite Scrolling Pull To Refresh Reduced Friction
		Engaging Strategy - Social Brokering	False Content Customization Regression Towards The Mean Social Connector
		Governing Strategy - Decision Uncertainty	Decision Uncertainty Clinging To Accounts Persuasive Language
		Governing Strategy - Labyrinthine Navigation	External Solution Search Labyrinth Hidden In Plain Sight
		Governing Strategy - Redirective Condition	Auto Accept Third Party Terms Decision Governing Forced Access Granting Forced Dialogue Interaction Forced Grace Period Plain Evil

Table 7.2: Used Dark Pattern Taxonomies

B Survey

Here is the english version of the online survey.

B.1 Social Media Usage and Habits

1. Please indicate which social media platforms you engage with on a daily basis by selecting them from the list below. (Multiple Choice Question)

   - Facebook
   - Instagram
   - TikTok
   - Snapchat
   - LinkedIn
   - BeReal
   - Reddit
   - YouTube
   - Twitter/X
   - Pinterest
   - Twitch
   - Other
   - None

Social Media User

1. On average, how much time do you spend on social media each day? (Single Choice Question)

   - 0 – 30 minutes
   - 30 minutes – 1 hour
   - 1 – 2 hours
   - 2 – 3 hours
   - 3 – 4 hours
   - 4 – 5 hours
   - 5+ hours

2. Please choose your main reasons for using social media. I use social media... (Multiple Choice Question)

- to know what's being talked about
- for entertainment and leisure
- to keep in touch with friends and family
- to communicate with (distance) friends
- to find potential partners (dating)
- because everyone else uses it
- to join groups related to my interests
- as a way to unwind and relax
- when no one else is available
- to expand my social circle and meet new people (networking)
- to read the news
- to share information that may be useful or interesting to others
- to escape from everyday life
- to learn about specific topics of interest

3. Have you ever read the privacy policy of any social media platform? (Single Choice Question)
   - Yes
   - Some parts of it
   - No
4. Have you ever changed your privacy settings on any social media platform? (Single Choice Question)
   - Yes
   - No
5. Have you ever used time tracking apps or features to control your daily social media usage? (Single Choice Question)
   - Yes
   - No
6. Have you ever tried to reduce the amount of time you spend on social media? (Single Choice Question)
   - Yes, and
     - I reduced the time drastically
     - I reduced the time a little bit

  - I didn't reduce the time
- No

7. Are the notifications for your social media accounts currently enabled? (Single Choice Question)
   - Yes
   - Occasionally / for certain apps
   - No

Non-Social Media User

1. Have you used social media in previous years? (Single Choice Question)
   - Yes
   - No

B.2 Social Media Problems

Social Media User

1. Please indicate how much you agree or disagree with each of these statements. (Likert-Scale: strongly disagree, disagree, somewhat disagree, neither agree, nor disagree, somewhat agree, agree, strongly agree)
   - I am satisfied with the amount of time I spend on social media
   - I usually spend more time on social media than initially intended
   - There are instances when I regret the time spent on social media
   - I sometimes find myself scrolling through social media instead of focusing on the task at hand
   - I find myself looking at my phone every time there is a notification
   - I frequently interrupt my ongoing activities to access social media
   - It would be difficult to live without social media
   - I consider myself as addicted to social media
2. Please indicate how much you agree or disagree with each of these statements. (Likert-Scale: strongly disagree, disagree, somewhat disagree, neither agree, nor disagree, somewhat agree, agree, strongly agree)
   - I filter the content I consume online
   - I sometimes find myself looking at meaningless content

- There are instances when I regret the time I spent watching random content on social media
- I want to have control over the information that social media uses for content personalization
- I believe I have control over the information that social media uses for content personalization
- I fact-check news articles or information I read on social media
- I believe that social media spreads fake news

3. Please indicate how much you agree or disagree with each of these statements. (Likert-Scale: strongly disagree, disagree, somewhat disagree, neither agree, nor disagree, somewhat agree, agree, strongly agree)

   - I perceive the content and ads I see on social media as an invasion into my privacy
   - I am concerned about my privacy and security on social media platforms
   - I am concerned about the amount of personal information social media companies possess about me
   - I believe I have control over the personal information collected by social media companies about me

4. Please indicate how much you agree or disagree with each of these statements. (Likert-Scale: strongly disagree, disagree, somewhat disagree, neither agree, nor disagree, somewhat agree, agree, strongly agree)

   - Social media has a positive impact on my mental well-being
   - Social media has a negative impact on my mental well-being
   - I compare myself to the people I am connected on social media
   - The number of friends/followers I have is a representation of people's approval of me
   - I envy others' lives based on the snippets they share on social media
   - Gaining new connections, likes, or positive comments on social media boosts my self-esteem
   - Losing connections or receiving only a few likes/comments on social media leads to a decrease in my self-esteem
   - Using social media creates a sense of belonging and connection with others
   - Using social media makes me feel lonely

5. Is there a particularly negative feature of social media you want to share? (Open-ended question)

6. If you could talk to a social media platform's managers, designers, and developers, what modifications or new features would you suggest? (Open-ended question)

7. Please indicate to which degree you perceive the following social media platforms as manipulative (Likert-Scale: not manipulative, somewhat manipulative, manipulative, strongly manipulative, don't know)

    - Facebook
    - Instagram
    - TikTok
    - Snapchat
    - LinkedIn
    - BeReal
    - Reddit
    - YouTube
    - Twitter/X
    - Pinterest
    - Twitch

Non-Social Media User - Never Used

1. What factors contribute to your decision not to use social media?

    - I have negative assumptions with social media
    - Reasons related to mental health
    - Concerns over data privacy
    - Concerns over data security
    - I just don't see the point of using it
    - Other

2. What modifications would encourage you to start using social media? (Open-ended question)

Non-Social Media User - Previously Used

1. Please select why you deleted your social media account(s)?

    - I rarely used it
    - I had negative experiences on the platform(s)

- Reasons related to mental health
- Concerns over data privacy
- Concerns over data security
- To reduce time spent on social media
- Other

2. What modifications would encourage you to reconsider using social media? (Open-ended question)

B.3 Dark Patterns in Social Media

Instruction: In this section, videos and pictures of social media elements are presented to you. Subsequently, pairs of contrasting adjectives are listed. The circles between the adjectives represent gradations between the opposites. Please express your agreement with the adjectives by ticking one circle per line that best aligns with your impression. If you find the feature unclear or believe that the adjectives do not describe the feature accurately, select the "don't know / not applicable" option. Audio is not needed.

I perceive this social media feature as...

- interesting ◯ ◯ ◯ ◯ ◯ ◯ ◯ not interesting
- organized ◯ ◯ ◯ ◯ ◯ ◯ ◯ cluttered
- supportive ◯ ◯ ◯ ◯ ◯ ◯ ◯ obstructive
- boring ◯ ◯ ◯ ◯ ◯ ◯ ◯ exciting
- unpredictable ◯ ◯ ◯ ◯ ◯ ◯ ◯ predictable
- revealed ◯ ◯ ◯ ◯ ◯ ◯ ◯ covert
- complicated ◯ ◯ ◯ ◯ ◯ ◯ ◯ easy
- inefficient ◯ ◯ ◯ ◯ ◯ ◯ ◯ efficient
- friendly ◯ ◯ ◯ ◯ ◯ ◯ ◯ unfriendly
- addictive ◯ ◯ ◯ ◯ ◯ ◯ ◯ non-addictive
- pressuring ◯ ◯ ◯ ◯ ◯ ◯ ◯ suggesting
- enjoyable ◯ ◯ ◯ ◯ ◯ ◯ ◯ annoying
- deceptive ◯ ◯ ◯ ◯ ◯ ◯ ◯ benevolent
- clear ◯ ◯ ◯ ◯ ◯ ◯ ◯ confusing

1. Please watch the following video, focusing on the popup and provide your rating below

2. Imagine you want to change your privacy settings and the settings are displayed as shown in the image.
   Please have a look at the picture and provide your rating below

3. Imagine you come across this information during the deletion of your social media account. Please have a look at the picture and provide your rating below

4. Imagine seeing the pre-defined settings of your privacy preferences as illustrated in the image. Please have a look at the picture and provide your rating below

5. Imagine being curious about why a particular ad is shown to you, prompting you to click the orange ?button. A pop-up appears, which among other things, offers the option to withdraw consent for targeted advertisement. After clicking on the mentioned hyperlink, you are redirected to a website explaining consent and how to manage it, without the ability to withdraw consent. Please have a look at the picture and provide your rating below

6. Please have a look at the picture, focusing on the two different options, and provide your rating below

7. Please have a look at the picture and provide your rating below

8. Please have a look at the picture and provide your rating below

9. Imagine you want to save a photo from social media and click on the relevant button. A pop-up appears, requesting access to your photos. Since your intention is solely to save a photo and not to upload one from your photo gallery to social media, you choose the *Select Photos...* option. Without selecting any photo, you proceed to click *Done*. This results in an error, causing the platform to fail to save the photo. This process is visualized in the image below. Please have a look at it and provide your rating below

10. Please have a look at the picture and provide your rating below

11. Please have a look at the picture and provide your rating below

12. Please have a look at the picture and provide your rating below

13. Social Media customizes the content shown to you. Please have a look at the picture and provide your rating below

14. Please watch the following video, focusing on the endlessly appearing content and provide your rating below

15. Social Media platforms incorporate the Pull To Refresh mechanism. When you swipe from the top to the middle, the content gets refreshed. This is demonstrated in the following video. Please watch it and provide your rating below

B.4 Demographics

1. How do you describe yourself?
   - Female
   - Male
   - Non-binary / third gender
   - Prefer to self-describe
   - Prefer not to say
2. How old are you? I am ... years old
3. In which country do you currently reside?
   - United States
   - Austria
   - Other Country
4. What best describes your employment status over the last three months?
   - Employed full-time
   - Employed part-time
   - Maternity leave/taking care of kids
   - Student
   - Retired
   - Unemployed
   - Other
5. Would you consider yourself technically savvy?
   - Yes
   - A bit
   - No

B.5 Feedback and Raffle

1. Do you have any feedback, comments or final remarks?
2. Would you like to enter the raffle to win an Amazon voucher? Your response will still remain anonymous
   - I would like to participate in the raffle. I agree that my email address will be saved until the winner is drawn. My answers will continue to be anonymous and my email address will not be passed on to third parties.

C Design

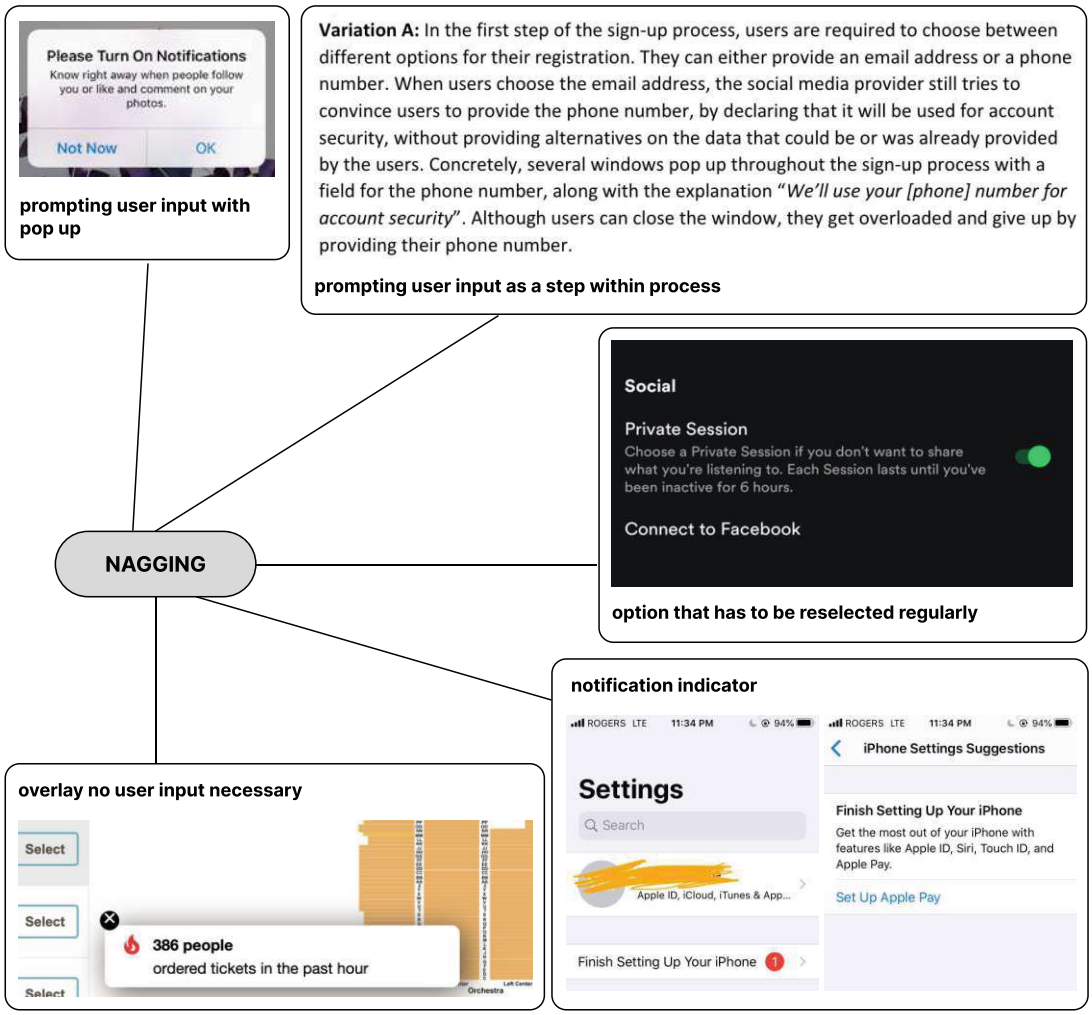


Figure 7.2: Mindmap Nagging Examples

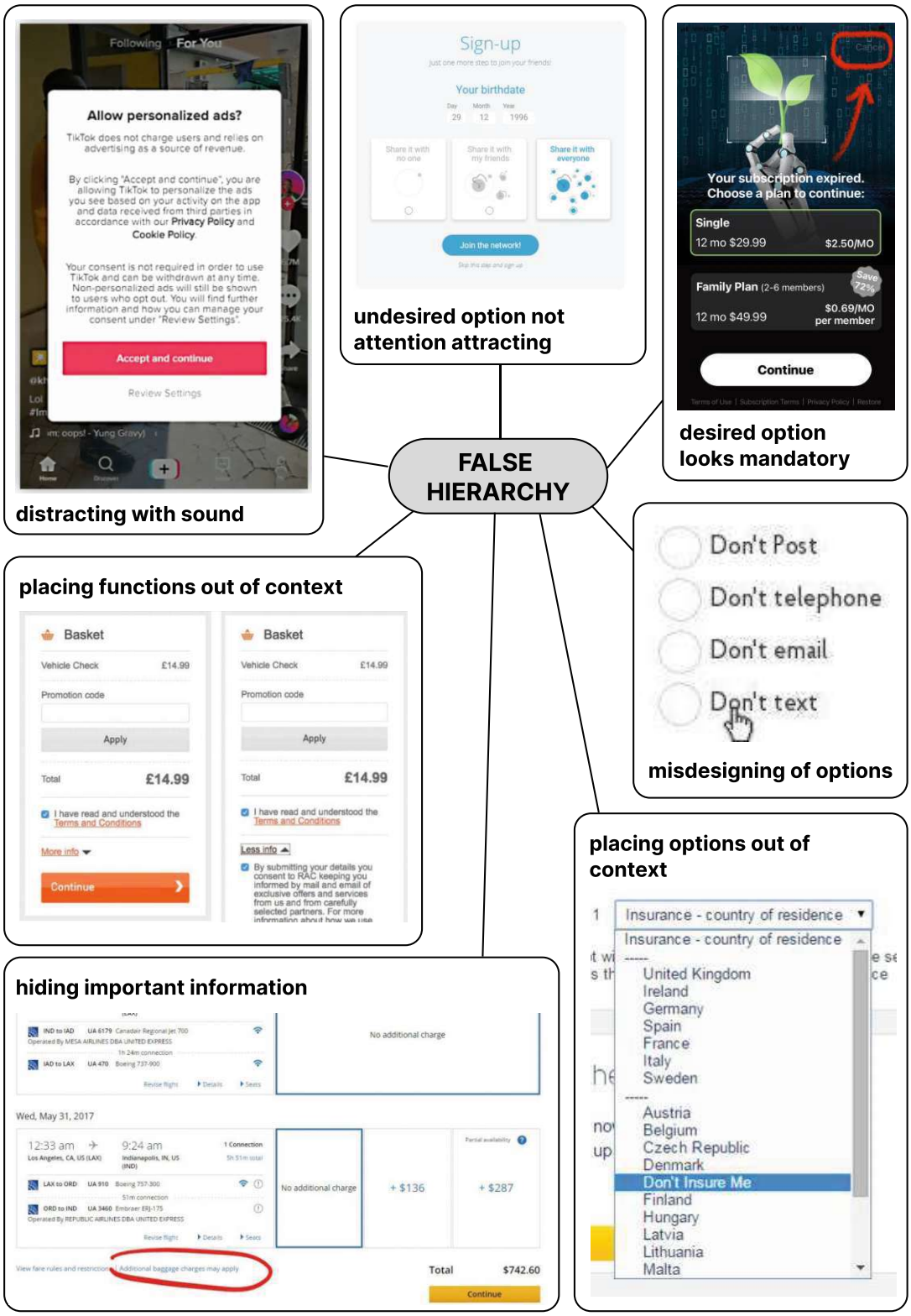


Figure 7.3: Mindmap False Hierarchy Examples

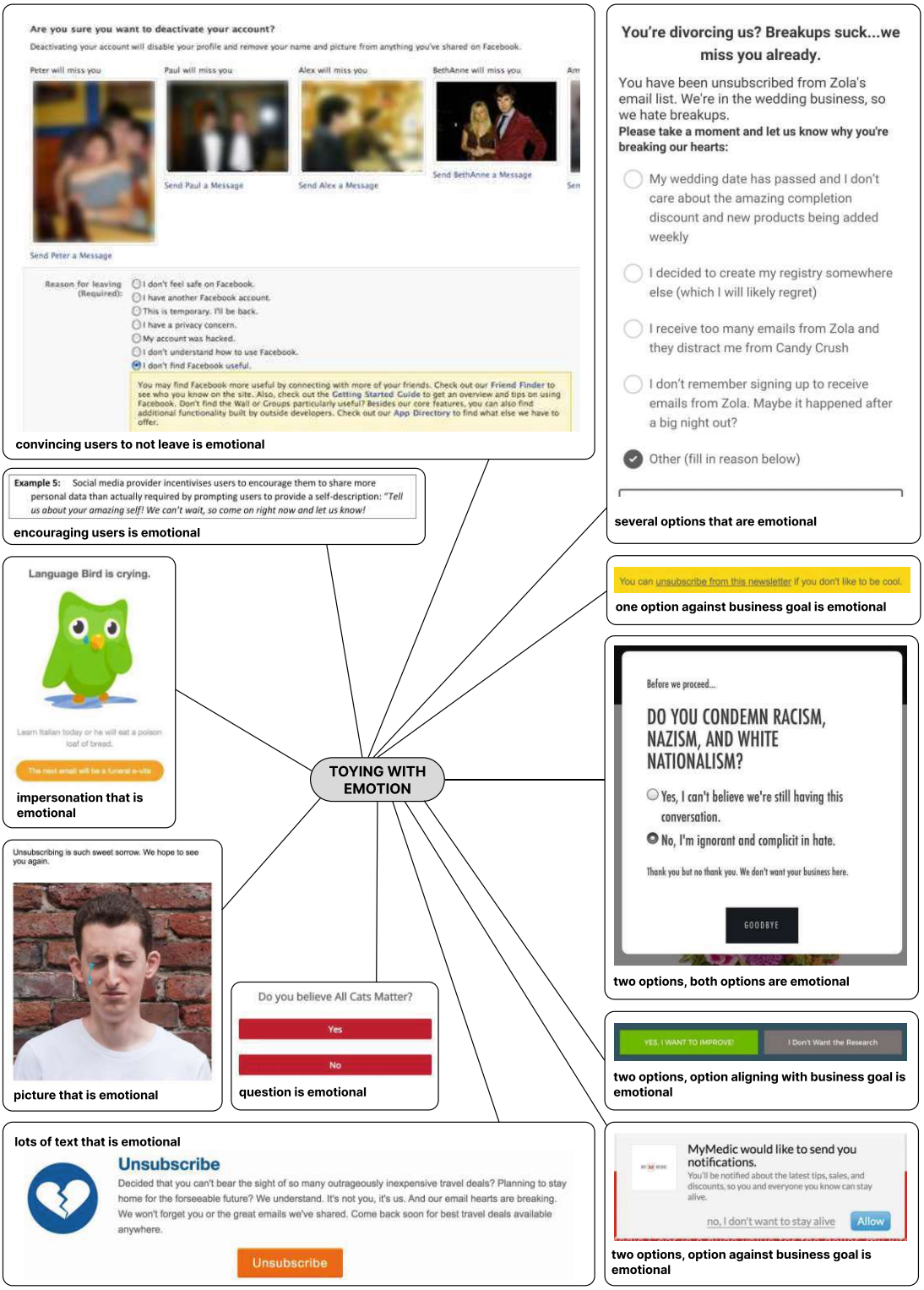


Figure 7.4: Mindmap Toying with Emotion Examples

Figure 7.5: Mash-up Ideation Method for Creating Bright Pattern Ideas: Columns represent survey participants' perceived social media problems, rows represent the dark patterns types used. Ideas incorporated into the feed are noted on a pink post-it, while ideas for the post or story creation process are written on a purple post-it.

D Assessment

D.1 Consent and Privacy Agreement

Consent and Privacy Agreement for User Study

1. Introduction
I appreciate your interest in participating in my user study as part of my thesis. Please read the following information carefully before consenting to participate in the study.

2. Study Details
Study Title: Evaluating Social Media Design Elements and Features
Investigator: Doris Maria Rhomberg
Institution/Organization: Vienna University of Technology in collaboration with researchers at Cornell Tech (NY, USA) and IU Bloomington (IN, USA)
Contact Information: e1649642@student.tuwien.ac.at

3. Purpose of the Study
The purpose of this study is to gather information about some social media features. Your participation will help me to understand user experiences and preferences.

4. Procedures
If you agree to participate in this study, you will be asked to evaluate social media features on your own in a survey setting and subsequently answer some questions in an interview setting. The study is expected to take approximately 45 minutes to complete.

5. Voluntary Participation
Your participation in this study is entirely voluntary. You may withdraw from the study at any time without any given reason. Should you decide to withdraw, any data collected from you will be destroyed and not used in the study.

6. Confidentiality and Data Protection
All data collected during this study will be treated confidentially. I will take appropriate measures to ensure that your personal data is protected.

Data Collected:
- Responses in the survey and answers to open questions in the interview
- Demographics, including age, gender, home country, and employment status
- Recordings of the study, including audio recordings, and possibly photos and notes taken during the conversation

Data Usage:
The data collected will be used solely for the purposes of my thesis and any resulting publications or presentations. Individual quotes may be used in publications, but they will be anonymous and not attributable to you.

8. Consent

By signing this consent form, you acknowledge that you have read and understood the information provided, and you agree to participate in this study.

Participant's Name: _____

Participant's Signature: _____

Date: _____

If you have any questions about this study or your rights as a participant, please contact me (Doris Rhomberg).

Thank you for your participation.

Figure 7.6: User Study Consent and Privacy Agreement

D.2 Interview Guide

Part 2 of User Study

Ask the following questions per pattern:

General
- What are your thoughts on this design element/feature?

Attractiveness *(How do users perceive the pattern?)*

- Did you feel manipulated by this design element/feature?
  - What made you feel like that?

Perceived Efficiency *(What perceptions do users hold regarding the effectiveness or influence of this pattern?)*

- What do you think are the intentions behind this design element/feature? Does the design element/feature address a specific user need or serve a practical purpose?
  - How effectively does this design element/feature fulfill its intended function?

- Would the design element affect your behavior in any way?
  - Like what would you do?

Acceptance *(When do users think such design manipulation is appropriate?)*

- Is it okay to feel manipulated in this case?

- Do you see something positive or negative about this design element/feature?
  - Do you have any preferences or concerns regarding this design element/feature?

Part 3

Ask the following questions once. Scroll videos to popup

Acceptance
- Is there one feature that stood out? Do you have a strong positive / negative opinion about any of these?

- I remember you said design element XYZ was manipulative. Do you excuse such manipulative behavior under certain circumstances?
  - Are there examples when such manipulative behavior is okay?
  - Are there examples when you desire such manipulative behavior?

- How do you characterize manipulation? What is manipulative for you? Is manipulation stronger if it's obvious or not?

Others
- Do you have any additional comments?

Figure 7.7: User Study Part 2 - Interview Guide

D.3 Thematic Analysis Codes

False Hierarchy Dark Pattern	False Hierarchy Bright Pattern
general like/dislike • #doesn't_like_it • #doesn't_like_feature • #like_to_have_a_choice good • #clear_instruction • #understandable • #not_manipulative_second_option bad • #forcing • #deceptive • #misleading #malicious • #privacy_concerning • #poor_ux • #annoying • #goals_don't_align_with_user's • #unequal_buttons • #not_manipulative_but • #manipulative • #encourage_bright_button	general like/dislike • #worst_one • #doesn't_like_feature • #likes_intention • #likes_it good • #clear_instruction • #interesting • #helpful • #beneficial • #thought_provoking • #intention_makes_sense • #not_manipulative • #not_manipulative_but • #encourage_bright_button bad • #annoying • #deceptive • #patronizing • #inefficient • #accessibility_concerning • #poor_ux • #deterring • #manipulative • #little_manipulative critique on implementation • #next_steps_unclear • #putting_work_on_users • #don't_want_to_read_papers • #users_research_by_themselves • #users_own_biases • #questioning_implementation • #other_stakeholders_should_fact_check • #not_for_personal_stuff • #unequal_buttons • #doesn't_like_censorship

Figure 7.8: Thematic Analysis Codes - False Hierarchy

Nagging Dark Pattern	Nagging Bright Pattern
general like/dislike • #worst_one • #doesn't_like_it • #doesn't_like_random_notifications • #likes_notification good • #understandable bad • #undermining_users_intelligence • #interrupting • #annoying • #angry • #cluttered • #frustrating • #demanding • #forcing • #undesired_options • #deceptive • #boring • #pestering • #manipulative • #not_manipulative_but	general like/dislike • #doesn't_like_feature • #likes_intention • #doesn't_like_it good • #useful #helpful • #appreciated_not_stopping • #taking_care_of • #cool • #thought_provoking • #funny bad • #disruptive, #moment_disruption • #cluttered • #annoying • #inducing_shame #judging • #frustrating • #confusing_popup • #patronizing • #privacy_concerning • #inefficient • #undesired_implementation • #poor_ux • #pressuring • #manipulative • #not_manipulative_but

Figure 7.9: Thematic Analysis Codes - Nagging

Toying with Emotion Dark Pattern	Toying with Emotion Bright Pattern
general like/dislike • #worst_one • #doesn't_like_it • #likes_it good • #thought_provoking • #interesting • #friendly • #not_manipulative in between • #sarcastically #humor bad • #pressuring • #inducing_guilt • #deceptive • #annoying • #disrespectful • #mocking • #emotional • #saddening • #inefficient #useless • #blackmailing • #undesired_language • #pathetic • #hate • #undermining_users_decision • #confusing • #no_choice_hate_friends_stay • #people_will_know_it_means_delete • #manipulative • #little_manipulative	general like/dislike • #favorite_one • #less_intrusive • #likes_it • #doesn't_like_it good • #helpful • #thought_provoking • #interesting • #friendly • #organized • #efficient • #relieving • #eye_catching • #not_manipulative in between • ironic bad • #annoying • #inducing_shame • #inducing_guilt • #patronizing • #confusing_pattern • #pressuring • #inefficient • #manipulative • #not_manipulative_but

Figure 7.10: Thematic Analysis Codes - Toying with Emotion

List of Figures

List of Tables

Acronyms

EDPB European Data Protection Board. 17, 31, 77

GDPR General Data Protection Regulation. 17, 37

SDS System Darkness Scale. 35

SUS System Usability Scale. 12, 18, 34, 35

UEEQ User Experience Ethics Questionnaire. 36–39, 42, 44–46, 65, 69–72, 80, 81, 103, 125

UEQ User Experience Questionnaire. 4, 18, 34–37, 42

UEQ-S User Experience Questionnaire Short. 36

Bibliography

[1] Brainstorm. https://www.designkit.org/methods/brainstorm.html. Last accessed: 2024-05-15.

[2] Ideation Method: Mash-Up. https://www.ideou.com/pages/ideation-method-mash-up. Last accessed: 2024-05-15.

[3] Mash-ups. https://www.designkit.org/methods/mash-ups.html. Last accessed: 2024-05-15.

[4] J. Aagaard, M. E. C. Knudsen, P. Bækgaard, and K. Doherty. A Game of Dark Patterns: Designing Healthy, Highly-Engaging Mobile Games. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI EA '22, pages 1–8, New York, NY, USA, Apr. 2022. Association for Computing Machinery.

[5] W. Adams. Conducting Semi-Structured Interviews. Aug. 2015.

[6] L. Alberts and M. Van Kleek. *Computers as Bad Social Actors: Dark Patterns and Anti-Patterns in Interfaces that Act Socially*. Feb. 2023.

[7] S. Avolicino, M. Di Gregorio, F. Palomba, M. Romano, M. Sebillo, and G. Vitiello. AI-Based Emotion Recognition to Study Users' Perception of Dark Patterns. In M. Kurosu, S. Yamamoto, H. Mori, M. M. Soares, E. Rosenzweig, A. Marcus, P.-L. P. Rau, D. Harris, and W.-C. Li, editors, *HCI International 2022 - Late Breaking Papers. Design, User Experience and Interaction*, volume 13516, pages 185–203. Springer International Publishing, Cham, Oct. 2022. Series Title: Lecture Notes in Computer Science.

[8] V. Bairagi and M. V. Munot, editors. *Research Methodology: A Practical and Scientific Approach*. Chapman and Hall/CRC, New York, Feb. 2019.

[9] S. Benartzi, J. Beshears, K. L. Milkman, C. R. Sunstein, R. H. Thaler, M. Shankar, W. Tucker-Ray, W. J. Congdon, and S. Galing. Should governments invest more in nudging? *Psychological Science*, 28(8):1041–1055, Mar. 2017. PMID: 28581899.

[10] K. Bongard-Blanchy, A. Rossi, S. Rivas, S. Doublet, V. Koenig, and G. Lenzini. "I am Definitely Manipulated, Even When I am Aware of it. It's Ridiculous!" - Dark

Patterns from the End-User Perspective. In *Designing Interactive Systems Conference 2021*, DIS '21, pages 763–776, New York, NY, USA, June 2021. Association for Computing Machinery.

[11] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, Jan. 2006.

[12] V. Braun and V. Clarke. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, Aug. 2019. Publisher: Routledge.

[13] H. Brignull. *Deceptive patterns: Exposing the tricks tech companies use to control you*. Testimonium Ltd, July 2023.

[14] H. Brignull, M. Leiser, C. Santos, and K. Doshi. Deceptive patterns – user interfaces designed to trick you, Apr. 2023.

[15] J. Brooke. SUS: A quick and dirty usability scale. *Usability Eval. Ind.*, 189, Nov. 1995.

[16] C. Bösch, B. Erb, F. Kargl, H. Kopp, and S. Pfattheicher. Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns. *Proceedings on Privacy Enhancing Technologies*, 2016:237–254, July 2016.

[17] D. S. Calawen. Dark Patterns: Effect on Overall User Experience and Site Revisitation. 2022. Publisher: Technological University Dublin.

[18] S. S. Chivukula, C. Watkins, L. McKay, and C. M. Gray. "Nothing Comes Before Profit": Asshole Design In the Wild. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, pages 1–6, New York, NY, USA, May 2019. Association for Computing Machinery.

[19] H. Cho, D. Choi, D. Kim, W. J. Kang, E. K. Choe, and S.-J. Lee. Reflect, not Regret: Understanding Regretful Smartphone Use with App Feature-Level Analysis. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):456:1–456:36, Oct. 2021.

[20] E. Commission, D.-G. for Justice, Consumers, F. Lupiáñez-Villanueva, A. Boluda, F. Bogliacino, G. Liva, L. Lechardoy, and T. Rodríguez de las Heras Ballell. *Behavioural study on unfair commercial practices in the digital environment – Dark patterns and manipulative personalisation – Final report*. Publications Office of the European Union, Apr. 2022.

[21] G. Conti and E. Sobiesk. Malicious interface design: exploiting the user. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 271–280, New York, NY, USA, Apr. 2010. Association for Computing Machinery.

[22] C. Courage and K. Baxter. Chapter 7 - interviews. In C. Courage and K. Baxter, editors, *Understanding Your Users*, Interactive Technologies, pages 246–311. Morgan Kaufmann, San Francisco, 2005.

[23] A. L. Cox, S. J. Gould, M. E. Cecchinato, I. Iacovides, and I. Renfree. Design Frictions for Mindful Interactions: The Case for Microboundaries. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 1389–1397, New York, NY, USA, May 2016. Association for Computing Machinery.

[24] L. F. Cranor. Cookie monster. *Communications of the ACM*, 65(7):30–32, June 2022.

[25] A. Curley, D. O'Sullivan, D. Gordon, B. Tierney, and I. Stavrakakis. The Design of a Framework for the Detection of Web-Based Dark Patterns. *Conference Papers*, July 2021.

[26] L. Di Geronimo, L. Braz, E. Fregnan, F. Palomba, and A. Bacchelli. 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*, CHI '20, pages 1–14, New York, NY, USA, Apr. 2020. Association for Computing Machinery.

[27] EDPB. Guidelines 03/2022 on deceptive design patterns in social media platform interfaces: how to recognise and avoid them. version 2.0 | European Data Protection Board, Feb. 2023.

[28] L. Ekroth and J. Sandqvist. *Confirmshaming and its effect on users : A qualitative study on how confirmshaming in unsubscription processes affect users*. June 2020.

[29] R. A. Elphinston, M. J. Gullo, and J. P. Connor. Validation of the Facebook Addiction Questionnaire. *Personality and Individual Differences*, 195:111619, Sept. 2022.

[30] W. L. i. R.-B. U. Experience. Affinity Diagramming: Collaboratively Sort UX Findings & Design Ideas. `https://www.nngroup.com/articles/affinity-diagram/`, Apr. 2024. Last accessed: 2024-07-12.

[31] M. Fansher, S. S. Chivukula, and C. M. Gray. #darkpatterns: UX Practitioner Conversations About Ethical Design. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, pages 1–6, New York, NY, USA, Apr. 2018. Association for Computing Machinery.

[32] B. J. Fogg. Introduction. In B. Fogg, editor, *Persuasive Technology*, Interactive Technologies, pages 1–13. Morgan Kaufmann, San Francisco, 2003.

[33] B. J. Fogg. Computers as Persuasive Social Actors. *Ubiquity*, Jan. 2011.

[34] C. Gray, S. Chivukula, J. Chen, S. Chivukula, and S. S. Chivukula. End User Accounts of Dark Patterns as Felt Manipulation. *Proceedings of the ACM on Human-Computer Interaction*, 5:Article 372, Oct. 2021.

[35] C. Gray, C. Santos, and N. Bielova. *Towards a Preliminary Ontology of Dark Patterns Knowledge*. Apr. 2023.

[36] C. M. Gray, Y. Kou, B. Battles, J. Hoggatt, and A. L. Toombs. The Dark (Patterns) Side of UX Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–14, Montreal QC Canada, Apr. 2018. ACM.

[37] P. Graßl, H. Schraffenberger, F. Zuiderveen Borgesius, and M. Buijzen. Dark and bright patterns in cookie consent requests. *Journal of Digital Social Research*, 3(1):1–38, Feb. 2021.

[38] S. Greenberg, S. Boring, J. Vermeulen, and J. Dostal. Dark patterns in proxemic interactions: a critical perspective. In *Proceedings of the 2014 conference on Designing interactive systems*, DIS '14, pages 523–532, New York, NY, USA, June 2014. Association for Computing Machinery.

[39] J. Gunawan, A. Pradeep, D. Choffnes, W. Hartzog, and C. Wilson. A comparative study of dark patterns across web and mobile modalities. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2), Oct. 2021.

[40] E. J. Johnson and D. G. Goldstein. Do Defaults Save Lives?, Nov. 2003.

[41] J. King and A. Stephan. Regulating Privacy Dark Patterns in Practice—Drawing Inspiration from California Privacy Rights Act, Sept. 2021.

[42] D. Kirkman, K. Vaniea, and D. W. Woods. DarkDialogs: Automated detection of 10 dark patterns on cookie dialogs. Apr. 2023.

[43] C. Kothari. *Research Methodology: Methods and Techniques*. New Age International (P) Limited, 2004.

[44] B. Laugwitz, T. Held, and M. Schrepp. Construction and Evaluation of a User Experience Questionnaire. In A. Holzinger, editor, *HCI and Usability for Education and Work*, Lecture Notes in Computer Science, pages 63–76, Berlin, Heidelberg, 2008. Springer.

[45] T. Liu, H. Wang, L. Li, X. Luo, F. Dong, Y. Guo, L. Wang, T. Bissyandé, and J. Klein. MadDroid: Characterizing and Detecting Devious Ad Contents for Android Apps. *Proceedings of The Web Conference 2020*, pages 1715–1726, Apr. 2020. Conference Name: WWW '20: The Web Conference 2020 ISBN: 9781450370233 Place: Taipei Taiwan Publisher: ACM.

[46] J. Luguri and L. J. Strahilevitz. Shining a Light on Dark Patterns. *Journal of Legal Analysis*, 13(1):43–109, Jan. 2021.

[47] K. Lukoff, A. Hiniker, C. M. Gray, A. Mathur, and S. S. Chivukula. What Can CHI Do About Dark Patterns? In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI EA '21, pages 1–6, New York, NY, USA, May 2021. Association for Computing Machinery.

[48] K. Lukoff, U. Lyngs, H. Zade, J. V. Liao, J. Choi, K. Fan, S. A. Munson, and A. Hiniker. How the Design of YouTube Influences User Sense of Agency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, pages 1–17, New York, NY, USA, May 2021. Association for Computing Machinery.

[49] A. M. Bhoot, M. A. Shinde, and W. P. Mishra. Towards the identification of dark patterns: An analysis based on end-user reactions. In *Proceedings of the 11th Indian Conference on Human-Computer Interaction*, IndiaHCI '20, page 24–33, New York, NY, USA, Nov. 2020. Association for Computing Machinery.

[50] M. Maier and R. Harr. Dark Design Patterns: An End-User Perspective. *Human Technology*, 16:170–199, Aug. 2020.

[51] S. Mansur, S. Salma, D. Awofisayo, and K. Moran. *AidUI: Toward Automated Recognition of Dark Patterns in User Interfaces*. Mar. 2023.

[52] A. Mathur, G. Acar, M. J. Friedman, E. Lucherini, J. Mayer, M. Chetty, and A. Narayanan. Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):81:1–81:32, Nov. 2019.

[53] A. Mathur, M. Kshirsagar, and J. Mayer. What Makes a Dark Pattern... Dark? Design Attributes, Normative Considerations, and Measurement Methods. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, pages 1–18, New York, NY, USA, May 2021. Association for Computing Machinery.

[54] T. Mildner, M. Freye, G.-L. Savino, P. R. Doyle, B. R. Cowan, and R. Malaka. Defending Against the Dark Arts: Recognising Dark Patterns in Social Media. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, DIS '23, pages 2362–2374, New York, NY, USA, July 2023. Association for Computing Machinery.

[55] T. Mildner and G.-L. Savino. Ethical User Interfaces: Exploring the Effects of Dark Patterns on Facebook. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI EA '21, pages 1–7, New York, NY, USA, May 2021. Association for Computing Machinery.

[56] T. Mildner, G.-L. Savino, P. R. Doyle, B. R. Cowan, and R. Malaka. About Engaging and Governing Strategies: A Thematic Analysis of Dark Patterns in Social Networking Services. In *Proceedings of the 2023 CHI Conference on Human*

Factors in Computing Systems, CHI '23, pages 1–15, New York, NY, USA, Apr. 2023. Association for Computing Machinery.

[57] A. Monge Roffarello and L. De Russis. Towards Understanding the Dark Patterns That Steal Our Attention. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI EA '22, pages 1–7, New York, NY, USA, Apr. 2022. Association for Computing Machinery.

[58] C. Montag, B. Lachmann, M. Herrlich, and K. Zweig. Addictive Features of Social Media/Messenger Platforms and Freemium Games against the Background of Psychological and Economic Theories. *International Journal of Environmental Research and Public Health*, 16(14):2612, Jan. 2019. Number: 14 Publisher: Multidisciplinary Digital Publishing Institute.

[59] C. Moser, S. Y. Schoenebeck, and P. Resnick. Impulse Buying: Design Practices and Consumer Needs. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–15, New York, NY, USA, May 2019. Association for Computing Machinery.

[60] C. Neyman. A Survey of Addictive Software Design. June 2017.

[61] S. Oberhofer, S. Overhage, and K. Kimppa. Dark Patterns of digital platforms - when are users fooled by nudging? Sept. 2022.

[62] D. Ozkaramanli, E. Özcan, and P. Desmet. Long-Term Goals or Immediate Desires? Introducing a Toolset for Designing with Self-Control Dilemmas. *The Design Journal*, 20(2):219–238, Mar. 2017.

[63] M. Potel-Saville. From dark to fair patterns? How can design-driven innovation and neurodesign help fighting against deceptive design. Mar. 2023.

[64] K. Proudfoot. Inductive/Deductive Hybrid Thematic Analysis in Mixed Methods Research. *Journal of Mixed Methods Research*, 17(3):308–326, July 2023. Publisher: SAGE Publications.

[65] A. K. Purohit, L. Barclay, and A. Holzer. Designing for Digital Detox: Making Social Media Less Addictive with Digital Nudges. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, pages 1–9, New York, NY, USA, Apr. 2020. Association for Computing Machinery.

[66] J. Redström. *Persuasive Design: Fringes and Foundations*, volume 3962. May 2006. Journal Abbreviation: Persuasive Technology Pages: 122 Publication Title: Persuasive Technology.

[67] H. Sandhaus. Promoting Bright Patterns. CHI '23 Workshop: Designing Technology and Policy Simultaneously, Apr. 2023.

[68] J. R. Saura, D. Palacios-Marqués, and A. Iturricha-Fernández. Ethical design in social media: Assessing the main performance measurements of user online behavior modification. *Journal of Business Research*, 129:271–281, May 2021.

[69] B. Schaffner, N. A. Lingareddy, and M. Chetty. Understanding Account Deletion and Relevant Dark Patterns on Social Media. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2):417:1–417:43, Nov. 2022.

[70] M. Schrepp. User experience questionnaire handbook. Sept. 2015.

[71] M. Schrepp, A. Hinderks, and J. Thomaschewski. Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence*, 4:103, Jan. 2017.

[72] R. Schäfer, P. M. Preuschoff, and J. Borchers. Investigating Visual Countermeasures Against Dark Patterns in User Interfaces. In *Proceedings of Mensch und Computer 2023*, MuC '23, pages 161–172, New York, NY, USA, Sept. 2023. Association for Computing Machinery.

[73] K. Shepherd. Brainstorming with Mash-Up Ideation. `https://linesof.com/2020/02/28/brainstorming-with-mash-up-ideation/`, Feb. 2020. Last accessed: 2024-05-15.

[74] D. Susser and V. Grimaldi. Measuring automated influence: Between empirical evidence and ethical values. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '21, page 242–253, New York, NY, USA, May 2021. Association for Computing Machinery.

[75] L. Sánchez Chamorro, K. Bongard-Blanchy, and V. Koenig. Ethical Tensions in UX Design Practice: Exploring the Fine Line Between Persuasion and Manipulation in Online Interfaces. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, DIS '23, pages 2408–2422, New York, NY, USA, July 2023. Association for Computing Machinery.

[76] A. Terpstra, P. Graßl, and H. Schraffenberger. Think before you click: how reflective patterns contribute to privacy. May 2024.

[77] R. H. Thaler and C. R. Sunstein. Nudge: Improving decisions about health, wealth, and happiness. *Constitutional Political Economy*, 19:356–360, Feb. 2008.

[78] H. Truong and A. Dalbard. *Bright Patterns as an Ethical Approach to Counteract Dark Patterns : A Closer Investigation of The Ethics of Persuasive Design.* June 2022.

[79] C. Utz, M. Degeling, S. Fahl, F. Schaub, and T. Holz. (un)informed consent: Studying gdpr consent notices in the field. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, CCS '19, page 973–990, New York, NY, USA, Oct. 2019. Association for Computing Machinery.

[80] C. van Nimwegen, K. Bergman, and A. Akdag. Shedding light on assessing Dark Patterns: Introducing the System Darkness Scale (SDS). pages 1–10. BCS Learning & Development, July 2022.

[81] Y. Wang, G. Norcie, S. Komanduri, A. Acquisti, P. G. Leon, and L. F. Cranor. "I regretted the minute I pressed share": a qualitative study of regrets on Facebook. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, SOUPS '11, pages 1–16, New York, NY, USA, July 2011. Association for Computing Machinery.

[82] J. Zagal, S. Björk, and C. Lewis. Dark patterns in the design of games. 2013.