

About Engaging and Governing Strategies: A Thematic Analysis of Dark Patterns in Social Networking Services

Thomas Mildner
University of Bremen
Bremen, Germany
mildner@uni-bremen.de

Gian-Luca Savino
University of St. Gallen
St.Gallen, Switzerland
gian-luca.savino@unisg.ch

Philip R. Doyle
University College Dublin
Dublin, Ireland
philip.doyle1@ucdconnect.ie

Benjamin R. Cowan
University College Dublin
Dublin, Ireland
benjamin.cowan@ucd.ie

Rainer Malaka
University of Bremen
Bremen, Germany
malaka@tzi.de

ABSTRACT

Research in HCI has shown a growing interest in unethical design practices across numerous domains, often referred to as "dark patterns". There is, however, a gap in related literature regarding social networking services (SNSs). In this context, studies emphasise a lack of users' self-determination regarding control over personal data and time spent on SNSs. We collected over 16 hours of screen recordings from Facebook's, Instagram's, TikTok's, and Twitter's mobile applications to understand how dark patterns manifest in these SNSs. For this task, we turned towards HCI experts to mitigate possible difficulties of non-expert participants in recognising dark patterns, as prior studies have noticed. Supported by the recordings, two authors of this paper conducted a thematic analysis based on previously described taxonomies, manually classifying the recorded material while delivering two key findings: We observed which instances occur in SNSs and identified two strategies — engaging and governing — with five dark patterns undiscovered before.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; HCI theory, concepts and models; *Empirical studies in interaction design*; Interaction design theory, concepts and paradigms; • **Security and privacy** → *Usability in security and privacy*.

KEYWORDS

SNS, social media, social networking services, interface design, dark patterns, well-being, ethical interfaces

ACM Reference Format:
Thomas Mildner, Gian-Luca Savino, Philip R. Doyle, Benjamin R. Cowan, and Rainer Malaka. 2023. 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), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3544548.3580695

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

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

© 2023 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-9421-5/23/04.
https://doi.org/10.1145/3544548.3580695

1 INTRODUCTION

"Reading this paper makes you a better person!" Emotionally pressuring language is just one example of a growing body of unethical and malicious design practices, often referred to as "dark patterns". Brignull [7] first introduced the term over a decade ago, describing dark patterns as interface design strategies that coerce or steer people into actions they would not necessarily engage in if fully informed [31]. Since then, significant effort has been expended on identifying, capturing, and describing examples of dark patterns, most notably in the domain of e-commerce websites where a large corpus of dark patterns has been catalogued [31]. There is, however, a need to extend this work further to capture dark patterns across other domains, such as social networking services (SNSs). Although SNSs are used extensively across the globe, a clear view as to the types of patterns used on these sites and how these vary between major SNSs platforms has yet to be defined. Existing literature does not focus exclusively on SNSs [4, 12, 15, 21] or focuses on specific elements only, for instance advertising dark patterns only [22] and SNSs' deletion processes [39]. The work does, however, emphasise an urgent need to deepen our understanding of how SNSs harm their users through unethical design practices [12, 21, 22, 39].

Gaining this understanding is a critical step in informing current efforts to legislate against such practices. Current legislation often limits regulation to data and privacy protection, in many cases neglecting potentially harmful consequences interface designs can have on individuals. For instance, both the GDPR [8] and CCPA [27] require website providers to make their reasons for collecting data transparently whilst also offering users the option to decline any storage of personal data. However, interface designs that encourage excessive use of social media or designs that obfuscate account deletion are currently unregulated. To support future regulatory endeavours, such as a proposed draft for guidelines by the European Data Protection Board (EDPB) [3], we first need to understand how SNSs use dark patterns and identify what domain-specific dark patterns might be used on these platforms.
Our work contributes to this understanding by answering two key research questions:

RQ1 What types of dark patterns are currently used in the four SNSs Facebook, Instagram, TikTok, and Twitter?
RQ2 Do SNSs contain dark patterns currently unique to their domain?

To address these questions, we had six HCI researchers conduct reviews across four widely used SNS platforms, based on their mobile applications: Facebook, Instagram, TikTok, and Twitter. Reviewers were assigned to identify instances of dark patterns whilst completing a range of tasks commonly carried out by users on these platforms. The decision fell on HCI researchers as the ability to not only recognise dark patterns but also to reflect and react to them was deemed a crucial requirement for reviewers to generate best possible results. Prior studies researching non-expert users' ability to recognise dark patterns demonstrate a significant difficulty for such tasks [4, 12] which we aim to mitigate by relying on experts instead. These sessions were recorded, producing 16 hours of interaction data, which was then evaluated using thematic analysis. Based on screen recordings created during the study, instances of 44 out of 80 previously established dark patterns were identified. Thematic analysis of these identified five consistent themes the describe SNS-specific dark patterns: (1) *interactive hooks*; (2) *social brokering*; (3) *decision uncertainty*; (4) *labyrinthine navigation*; and (5) *redirective conditions*. These themes were then further organised into two overarching strategies that cover more high-level incentives allowing for broader application: engaging strategies and governing strategies. These emerge from SNS-specific incentives that differ from how dark patterns are used in other domains. The strategies include two and three types of the SNS-specific dark patterns. Falling under the umbrella of engaging strategies are the *interactive hooks* and *social brokering* dark patterns, which are designed to keep users occupied and entertained with SNS for as long as possible. Whereas, SNS dark patterns that can be considered governing strategies include *decision uncertainty*, *labyrinthine navigation*, and *redirective conditions*, which are designed to navigate users' decision-making ability on these platforms. We contribute to the current dark pattern discourse and literature by considering the impact dark patterns have on four popular SNSs, and by extending current taxonomies with instances specific to this domain.

2 RELATED WORK

In this section, we will review the relationship between research on dark patterns and work on the persuasive design in social media. We highlight the necessity to carry the dark pattern discourse over to SNSs, bridging the current gap between these two strands of research. The first two subsections present a brief overview of current dark pattern taxonomies, which were used to guide the thematic analysis conducted in this study. The overview also highlights approaches and methodologies used in previous work aimed at identifying dark patterns. Some of these are adopted in this study which builds on previous work to better understand dark patterns, specifically within the context of SNS interface design. We then continue with work that dealt with users' perception of dark patterns and their ability to recognise them in different environments. Lastly, we establish the importance of considering dark patterns in social media while following the discourse of SNS interface strategies that lead to problematic or even harmful usage behaviour.

2.1 Early Research On Dark Patterns

The past decade of research into dark patterns has defined and described a comprehensive taxonomy of different types across several different domains. In recent work, Mathur et al. [32] offer a summary of the current dark pattern landscape resulting in a dense taxonomy comprising relevant works. This taxonomy is the result of their attempt to characterise dark patterns based on the cognitive biases that they exploit. As this corpus depicts a thorough overview of past dark pattern collections, we decided to use this corpus as the basis for our study. However, we decided only to include work based on empirical research, which means that we excluded reports, such as the NCC [10] and the CNIL [14] that promote dark patterns to the public while recommending guidelines to make informed decisions.

One of the earliest pieces of research on dark patterns was conducted by Brignull [7], who captured and described interface strategies used to harm people through interface tricks. Producing the first taxonomy of dark pattern types, Brignull described twelve interface tricks designed to misguide users. Among these were dark patterns, such as *sneak into basket*, *hidden costs*, and *Price Comparison Prevention*, which all operate by obscuring certain information from users of online shopping sites. The intention here is to inhibit customers' ability to make informed decisions and potentially mislead them into buying unwanted products. Alternately, the dark patterns *forced continuity*, *privacy zuckering*, and *roach motel* use strategies that limit the options and decisions available to people when using online services. In the same year as, Brignull, Conti and Sobiesk introduced their own taxonomy based on findings from a twelve-month-long study aimed at cataloguing a wide range of malicious interface practices. This second taxonomy includes eleven types of dark patterns, such as *coercion*, which describes interfaces that mandate users' decisions by restricting alternative options and enforcing compliance. Other techniques noted by the authors include *interruptions* that interfere with a user's task flow and the *obfuscation* of important information, both of which operate by hindering informed decision-making. Elsewhere, Zagal et al. examined dark patterns in video games, identifying seven types of dark patterns that specifically focus on game mechanics [47]. The research shows that while certain patterns exploit a game's ecosystem of connected users, such as *social pyramid schemes* and *impersonation*, others impact game-play experience like *grinding* and *playing by appointment*.

Elsewhere, Greenberg et al. [20] consider dark patterns in conjunction with proxemics theory [23]. Identifying nine types of dark patterns in total, the authors discuss interactions with potentially abusive systems in spatial environments. For example, the *attention grabber* and *disguised data collection* dark patterns could be used in the design of digital billboards and involves brands exploiting people's proximity and personal data to deliver personalised advertising to specific pedestrians as they pass by. In a similar vein – inspired by the concept of *Privacy by Design* project [24] – Bösch et al. introduce nine further types of dark patterns, which are effectively inverse strategies to the privacy strategies developed in the Privacy by Design project. The work also highlights the role design strategies can play in manipulating users both for good and for nefarious reasons. Collectively, these early studies show that dark patterns can appear in a variety of contexts and situations, highlighting the importance of establishing a broad understanding of their origins.

2.2 Understanding the Origins of Dark Patterns

Reflecting on Brignull's original work [7], Gray et al. looked to investigate how dark patterns are created in the first place. Here, researchers adopted a qualitative approach, using established taxonomies to analyse an image-based corpus of potential types of dark patterns [19]. The work defines five high-level strategies that practitioners engage in when developing manipulative designs. For instance, the *obstruction* dark pattern was used to make processes unnecessarily difficult, and incorporates Brignull's *roach motel*, *price comparison prevention*, and *intermediate currency* [7]. In later work that adopts a similar approach, Gray et al. analysed 4775 user-generated posts of the Reddit sub-forum *r/assholedesign* [17]. Following multiple iterations of content analysis, the authors describe a set of six properties of "asshole designers" that portray malicious motivations of designers. The *two-faced* property, for instance, describes designers who offer conflicting information that limits users' ability to make an informed decision. Understanding the origins of dark patterns as constraints under which practitioners work offers relevant insights into where to look when trying to recognise dark patterns anywhere.

Focusing more centrally on the frequency with which dark patterns are embedded in online interfaces, Mathur et al. [31] built a web-crawler application to collect data from over 11K shopping websites. The work identifies instances of dark patterns in more than 11% of their samples. Although this percentile already presents a significant number of occurrences, the authors limited the breadth of their analysis by not analysing imagery material and suggest that many more dark patterns could be identified had other factors been taken into account. Nonetheless, guided by prior works from Gray et al. [19] and Brignull [7], the authors were able to compose a taxonomy of fifteen types of dark patterns. Extending these works further, Mathur et al. [32] more recently looked to identify clusters or relationships among established dark patterns, setting the basis for this work's thematic analysis. Taking the existing five high-level characteristics from their previous work [31], the authors added a sixth characteristic that incorporates Zagal et al.'s online gaming dark pattern *disparate treatment* [47]. Their proposed model further categorises these six dark pattern characteristics into two choice architectures that distinctly affect users: (1) modification of decision space and (2) manipulation of information flow. The authors thus propose an interesting three-tiered hierarchical framework (choice architectures > high-level dark pattern characteristics > specific manifestations of dark pattern designs) under which discovered and yet-to-be-defined dark patterns can fit into. Overall, the above-mentioned work collectively describe 81 specific types of dark patterns from various domains, mostly unique in how they operate. However, we noticed an omission in regard to SNSs. Filling this gap, we build on previous work by developing a deductive codebook containing this taxonomy, which was then used during the execution of the thematic analysis.

2.3 Recognising and Identifying Dark Patterns

With a more central focus on end-users' perspectives of dark patterns, Di Geronimo et al. [12] inspected popular mobile applications sampled from the Google Play Store. The authors use a cognitive walkthrough [37] to identify dark patterns across a total of 240 apps, each used for ten minutes. Findings showed that 95% of the tested applications contained dark patterns. In a second study, the authors evaluate users' ability to recognise dark patterns. Using an online survey, the work suggests most users had problems recognising dark patterns. Adopting a similar user-centred approach, Bongard-Blanchy et al. study peoples' awareness of manipulative interface designs and their recognition of dark patterns [4]. The study, which consisted of an online survey of 413 participants, also found that 59% of their participants were able to identify "interface elements that can influence users' choices" more than half of the time, suggesting that they were somewhat able to recognise dark patterns. This aligns with previous results from Di Geronimo et al. [12] and Maier and Harr [30]. However, the work also showed that while participants understood what dark patterns were and how they might manifest, they were still deceived by them during interactions. Elsewhere, Gunawan et al. [21] use thematic analysis based on video recordings of online services to study differences between the various web modalities and resulting dark patterns. A contribution of their analysis is the addition of a further twelve specific types of dark patterns to the body of work. Their *account deletion roadblocks* dark pattern, for instance, describes the insufficient communication between the service provider and user when the latter is trying to delete their account. Together, these works demonstrate processes for assessing and evaluating dark patterns and how they manifest. In this work, we consider a wider corpus of dark patterns on a relatively small set of applications. This allows us to offer profound insights in SNS-specific dark patterns while building on established work and their methodologies.

2.4 Design Strategies on SNSs

Social media plays a major role in the daily lives of millions of people worldwide. Although prior dark pattern work has considered SNS in their research [12, 21, 22], we still lack a thorough understanding of social media-specific dark patterns and the different kinds of harm they might present compared to malicious practices observed in other domains. In this context, early work by Roffarello and De Russis [35] proposes five dark patterns in Facebook and YouTube which aim to capture their users' attention. Although offering insights into SNS specific strategies, dark patterns that exploit alternative strategies than attention-capturing have yet to be described. Aside from work with an explicit focus on dark patterns, a growing body of work suggests certain uses and users of SNS may be prone to negative consequences in terms of mental health and well-being [1, 41, 43, 45]. A well-documented reason for this may be found in research comparing users' self-reported time spent on SNS to actual times, suggesting a lack of self-control and self-determination. Indeed, numerous studies have demonstrated a lack in users' ability to self-report accurately the time they spent on any SNS [13, 26, 40]. By showing that users generally spend less time on Facebook than they think while opening the application more often than realised, both Junco's and Ernala et al.'s works highlight a problem around self-awareness when it comes to frequency and length of social media use. A similar disparity was described by Mildner and Savino [34], where a contradiction in people's perceived and actual usage behaviour was observed among 116 participants. Most Facebook users who participated in

the study admitted that they spent more time on Facebook than planned, though most also declared they had no desire to spend less time on the platform. Another interesting result of their survey is noted in participants' perceived feeling of low control over ad-related data. These findings are in line with prior research noting a general dissatisfaction in users seems to arise from limited options to protect their privacy combined with an urge to have more control over their personal data [11, 44]. Investigating advertising controls on Facebook, Habib et al. [22] consider the impacts of dark patterns when conducting an online survey to support a thematic analysis identifying users' desired advertisement controls. The authors described users' difficulty in finding Facebook's Ad Preference section, to begin with. These findings affirm a prior suggestion by Gunawan et al. [21], who argue that the granularity of interfaces may discourage users from making desired changes to their preferences. Both users' difficulty in self-reporting the amount of time spent on SNSs and their lack of agency to control personal settings outline unethical practices that may fall under the umbrella of dark patterns. Consequently, we see an urgent need to better understand dark patterns in SNSs considering domain-specific strategies. This necessity is further highlighted by Schaffner et al. [39], who demonstrate how difficult account deletion is across 20 popular SNSs. Not only does the possibility to entirely delete an account vary depending on the modality of a particular service, but the authors further notice a difficulty among users to follow through with a deletion process. In our research, we further investigate this problem by recording and analysing reviewers' usage of four SNSs, a deletion process, allowing for detailed insights.

3 EXPERT REVIEW & DATA COLLECTION

To collect necessary data of SNS usage for the thematic analysis, we asked six HCI experts to record their usage of four mobile SNS applications. By expanding the scope, we gain a general understanding of common practices across SNSs. The decision fell on Facebook, Instagram, TikTok, and Twitter, as they present some of the most popular SNSs while satisfying comparable user needs. To assist their review in targeting potential unethical practices, each expert reviewer was provided ten tasks that afforded them to use the mobile applications of each SNS intensively. To get a better understanding of their actions, we also asked them to narrate their decisions to retrieve data in the form of a think-aloud protocol [25]. As we estimated that 30 minutes were required to complete all tasks, we asked each reviewer to record their usage of two of the four SNS. Thus, three independent recordings per SNS were collected. As the experiment was conducted during the COVID-19 pandemic, reviewers completed the study without supervision.

3.1 Reviewers

For reviewers, we reached out to HCI researchers, choosing six to investigate the four SNSs (3 female, 3 male, mean age = 28.33 years, SD = 1.63). All have multiple years of experience in HCI research, with backgrounds in cognitive science, computer science, and media science. At the time of the study, their years of experience varied from two to six years, with a mean of 3.83 years (SD = 1.47). At the time of conducting this research, all reviewers were employed as researchers at academic research faculties based in Germany, focusing on HCI related topics. Five reviewers are German citizens, while one is of Russian nationality. Except for one reviewer, participants did not have prior experience conducting dark pattern related research although all shared a general conceptualisation of the field. Participation in this study was voluntary and without compensation. The decision fell on HCI researchers to conduct this study as regular users have been repeatedly shown to have difficulties in detecting dark patterns [4, 12]. In contrast, the researcher's strong expertise of usability best practices and design heuristics makes them more sensitive towards interface strategies, enabling them to uncover and discuss persuasive techniques better. Hence, their expertise allows them to identify a wider variety of dark patterns than regular users would be able to. Nonetheless, we acknowledge that this qualitative research was conducted entirely by people with strong HCI backgrounds, possibly influencing our findings and interpretations.

3.2 Preparation

To counter any systematic issues caused by a particular operating system, each reviewer was provided with two smartphone devices: an iPhone X with iOS and an Android running Google Pixel 2 device. Reviewers were asked to use both devices simultaneously and to pay attention to device-specific differences when solving the tasks. Before giving them the devices, each phone was factory reset and only contained fresh installations of the SNS[1] and some media content containing photos with creative-commons licenses. Each participant was also provided with a new phone number and email address, allowing them to create a new social media account while protecting their privacy. For the same reason, the media content was provided, as some of the tasks required the participants to create and post content.

3.3 Tasks

Each evaluator was asked to execute ten tasks during their recording sessions. We decided to implement five tasks similar to Di Geronimo et al.'s [12] tasks as they were shown to be effective for this kind of study. To ensure coverage of SNS-specific interface problems, we also added five additional tasks tailored to this domain. Before employing the tasks to the reviewers, they were tested to ensure their accuracy and ability to cover a wide range of SNS-specific functionalities. After minor revisions following piloting, the ten tasks comprised the following exercises (whereas tasks that were taken from or worded closely to Di Geronimo et al.'s experiment are highlighted by an asterisk):

1. *Turn on screen recording on each device.*
*2. *Open the app and create an account to log in and then out.*
*3. *Close and reopen the app.*
4. *Create any kind of content, post it, and delete it.*
5. *Follow and unfollow other accounts.*
*6. *Visit the personal settings.*
*7. *Visit the ad-related settings.*
*8. *Use the application for its intended use (minimum of five minutes):*

[1]Installed versions consistent throughout the study: Facebook (iOS: 321.0.0.53.119; Android: 321.0.0.37.119); Instagram: (iOS: 191.0.0.25.122; Android: 191.1.0.41.124); TikTok (iOS:19.3.0; Android: 19.3.4); Twitter (iOS: 8.69.2; Android: 8.95.0-release.00).

I *Describe the natural flow of the app – what did you use it for?*
II *Could you use the app as you wanted, or did some features 'guide' your interactions?*
III *how easy was it to get distracted, and if so, what distracted you?*

9. *Delete your account.*
10. *Turn off screen recording and save the recording.*

3.4 Procedure

Before completing their walkthrough of the SNSs, reviewers were prepared for the study via a one-hour introduction session, outlining recent research on dark patterns and established taxonomies. This took the form of an online meeting. The introduction involved instructions around the 81 dark pattern types provided in the related work section, which was followed by an open discussion aimed at answering any unresolved questions reviewers may have. Afterwards, each reviewer was provided with a copy of the information, including a detailed list of dark patterns described in the introduction session. This material is included in the supplementary material of this work. Even though all participants have backgrounds in the field, this preparation ensured a common understanding of current conceptualizations of dark patterns. Information provided to reviewers also included a sheet explaining the cognitive biases particular dark patterns were designed to exploit [31]. This was included to help them identify dark patterns that were not already described in taxonomies. Each reviewer was then randomly assigned two of the four SNSs, ensuring that each SNS was examined independently three times by three different HCI researchers. The SNS applications that reviewers used were randomly assigned while ensuring that collectively reviewers did every possible sequential permutation.

Reviewers then conducted their reviews without supervision, following a detailed step-by-step guide they were provided with. The steps can be seen in Figure 1, and explain each task along with when to start and end the screen and audio recording. The manual further reminded reviewers to openly discuss their actions allowing us to retrieve data in the form of a think-aloud protocol, which we later evaluated to understand their decisions. The quality of the reviews are twofold: (1) The tasks were designed to get detailed insights into SNSs, affording reviewers to engage with the available features of each platform. (2) By recording the reviewers' commentary, we also gained information about their perception and judgement of potentially malicious artifacts. All recordings were automatically stored on each device and retrieved when reviewers returned both smartphones.

4 FINDINGS

As a result of this data collection, the recordings provided over 16 hours of video and audio material, including the think-aloud commentary from reviewers. On average, each session lasted 41 minutes ($SD = 19$) for each device. Between the six participants, each SNS was thus used for an average of 247 minutes. Looking at different approaches to identifying dark patterns, we noticed varying methodologies in prior work. Of the eight taxonomies considered here, two relied on crowd-sourced or user-generated imagery material to apply qualitative methodologies, such as the constant comparative method [17, 19], whereas another used a web crawler to generate a large enough corpus to apply hierarchical clustering [31]. Although highlighting important interface issues that need to be addressed, the remaining five works provide limited information to recreate their results. We decided to follow prior works and conducted a thematic analysis based on the 81 dark pattern types contained in this literature while further relying on screen recordings and audio commentary of participants. Furthermore, by limiting the scope of this research to four applications only, we were able to study each service more extensively.

4.1 Analysis of the Data

To answer our research questions, two researchers conducted a thematic analysis supported by the resulting data from the screen and audio recordings [6]. While the screen recordings contained the visual captures of dark patterns and interactions thereof, the audio served as complementary material, which allowed the coders to get a detailed impression of reviewers' perception and judgement of a scene. This aided the coding process in a supportive role. However, all labelling was conducted by the two coders. The thematic analysis was carried out through a combination of deductive and inductive coding [33] using the software ATLAS.ti [16]. In line with works from Di Geronimo et al. [12] and Gunawan et al. [21], we applied codes whenever an interface was perceived problematic, rather than being driven by a concern with designer intent. This allows for the identification of dark patterns that emerge from both intentionally manipulative and unintended yet potentially harmful design choices. The thematic analysis was conducted by two researchers who designed and administered the experiment and aimed to identify indications of dark patterns present in the four SNSs. Although the recorded sessions were coded independently, both coders met for discussions after each session to align their interpretations of dark pattern definitions, rendering testing for inter-rater reliability unnecessary [2]. Additionally, all instances where coders diverged from reviewers' comments were specifically marked and later discussed among both coders. Once all sessions were analysed, they met for a final thorough discussion to establish full agreement over all 12 sessions.

4.1.1 Creating The Deductive and Inductive Codebook. The deductive codebook was derived from the descriptions of existing dark patterns as seen in Table 1. Each of the 81 codes included in the codebook denotes a specific type of dark pattern. That is 81 specific interface design elements that - by accident or by design - impinge on the users' autonomy. As the 12 dark patterns by Gunawan et al. were not described by the time we conducted the thematic analysis, we were not able to include them in this study. In an attempt to reduce the number of codes, we collapsed dark pattern types that shared the same name. This resulted in combining Brignull and Mathur et al.'s *confirmshaming* dark patterns, which share a name and very similar descriptions, which reduced the codebook to 80 codes from an initial 81. Other candidates were also considered (*privacy zuckering* [5, 7] and *bait and switch* [7, 20]), but were not collapsed as their descriptions were deemed too distinct. We thus concluded the deductive codebook with 80 codes.

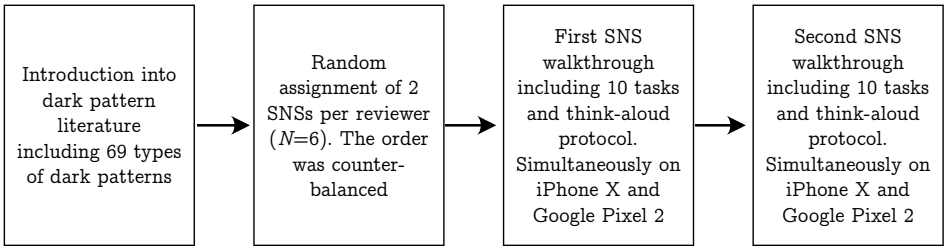


Figure 1: This flowchart describes the four steps in which our data was collected. First, reviewers received an in-depth introduction into the dark pattern literature. Each reviewer was then randomly assigned two of the four SNSs Facebook, Instagram, TikTok, and Twitter. Per SNS, the reviewers conducted a walkthrough based on 10 tasks on two devices (iOS and Android).

Once the deductive codebook had been generated, a random session was chosen and then independently coded by both coders. Deductive codes were applied wherever a type of dark pattern from the taxonomy sufficed to describe a recognised problem within the interface. In those cases where an aspect of the interface was deemed a potential dark pattern but was not present within the deductive codebook, a new code was generated and added to a separate inductive codebook. Descriptions for inductive codes focused on describing the design or interaction extracted from the recordings. Once the first session was completed, the two annotating researchers met to discuss the adequacy of established inductive codes and to form a common agreement. Afterwards, overlapping codes were eventually merged to compile a single inductive codebook comprising 22 codings by using the affinity diagramming technique [2].

4.1.2 Coding Remaining Data. Both coders then proceeded to code the remaining sessions independently, following the same procedures, only interrupted to discuss and resolve inconsistencies after each session, based on Blandford et al. [2]. By relying on screen recordings of the interfaces, instead of sampled screenshots, we gained an important advantage that allowed us to thoroughly investigate not only dark patterns on particular frames but observe sequenced interactions that would be invisible in still images. This was further affirmed by the recordings of the think-aloud protocol. If necessary, we could stop, repeat, and compare interface situations to get a deep understanding of the interactions the reviewers performed. Because of this decision, we also noticed dark patterns that only surfaced after sequential interactions were taken. This allowed us to observe dark patterns that were not previously described and were potentially unique to SNSs.

4.2 Deductive Codebook Analysis: Findings

Answering our first research question, we applied 44 out of the 80 deductive codes across our dataset, highlighting how many distinct dark patterns occurred across the four different SNSs. Table 1 displays all dark patterns and shows for which SNS they were applied during the analysis. Of these 44 that were applied, we observed 32 instances on each of the four SNSs. Regarding the original scope of individual dark pattern taxonomies, we noticed that the dark patterns described by Conti and Sobiesk [9] and Gray et al. [19] were noticed most commonly in the context of SNS, suggesting easier applicability of these taxonomies. Interestingly, these sets of dark patterns were both originally created by analysing a wide range of interfaces rather than focusing on a specific domain. Further, those described by Gray et al. [19] even subsumed dark pattern types formerly described by Brignull [7], generalising them even further. We also investigated which SNSs feature the largest variety of dark pattern types. Facebook, which contained 41 different types of dark patterns, exhibited the most variety, followed by Instagram, featuring 39, Twitter, where 35 different types of dark patterns were observed, and finally Tiktok, with 37 different types identified.

While this result highlights that SNS seem to make use of a wide variety of dark pattern types, 36 dark pattern codes were still left unused, implying that these were not appropriate to describe dark patterns in SNSs. Three groups of dark patterns from the initial 80 were not or only rarely applied: the proxemic dark patterns by Greenberg et al. [20], the gaming dark patterns by Zagal et al. [47], and the e-commerce dark patterns by Mathur et al. [31]. Many of these dark patterns were generated by describing design decisions in a specific domain or context, making it almost impossible to apply them elsewhere. One example would be Greenberg's *captive audience* dark pattern, which requires a "person [to] [enter] a particular area" [20] relying heavily on the actual proximity, hence why they are less applicable to a domain as ubiquitous as SNSs.

Beyond overly-specific dark patterns, others share a name, albeit with varying definitions. As a result, two dark patterns with the same name do not necessarily apply identically to SNS. For example, *privacy zuckering* was described by Brignull [7] and Bösch et al. [5]. While Bösch et al. focuses specifically on sharing more data than a user intends to give based on privacy settings, Brignull does not make this specification and thus leaves more room for interpretation and application. Hence, Brignull's version of the dark pattern was applied across all four SNS during our analysis, while the version by Bösch et al. was only applied explicitly on Facebook and TikTok. In other cases, some names of dark pattern promise applicability in SNSs but were actually never applied as their definition hindered accurate coding. An example is illustrated by the dark pattern *immortal accounts* by Bösch et al. [5]. While the name of the dark pattern suggests that user accounts are (almost) impossible to delete, its definition refers to service providers requiring new users to sign up for accounts to use their service.

4.3 Inductive Codebook Analysis: Findings

Although 44 codes from the deductive codebook could be applied in the SNSs, often, new codes were required to describe interface issues

Author	Dark Pattern	F	I	Ti	Tw
Brignull [7]	Bait And Switch	●	○	○	●
	Confirmshaming	●	●	●	●
	Disguised Ads	○	●	○	●
	Forced Continuity	○	○	○	○
	Friend Spam	○	○	○	○
	Hidden Costs	○	○	○	○
	Misdirection	●	●	●	○
	Price Comparison Prevention	○	○	○	○
	Privacy Zuckering	●	●	●	●
	Roach Motel	●	●	●	●
	Sneak Into Basket	○	○	○	○
	Trick Question	○	○	○	○
Conti & Sobiesk [9]	Coercion	○	○	○	○
	Confusion	●	○	○	●
	Distraction	●	●	●	●
	Exploiting Errors	○	○	○	○
	Forced Work	●	●	●	●
	Interruption	●	●	●	●
	Manipulating Navigation	●	●	●	●
	Obfuscation	●	●	●	●
	Restricting Functionalities	●	●	○	○
	Shock	○	●	○	○
	Trick	○	○	○	○
Zagal et al. [47]	Grinding	○	○	○	○
	Impersonation	○	○	○	○
	Monetized Rivalries	○	○	○	○
	Pay To Skip	○	○	○	○
	Playing By Appointment	○	○	○	○
	Pre-Defined Content	○	○	○	○
	Social Pyramid Schemes	○	○	○	○
Greenberg et al. [20]	Attention Grabber	●	●	●	○
	Bait And Switch	○	○	○	○
	Captive Audience	○	○	○	○
	Disguised Data Collection	○	○	○	○
	Making Personal Info. Public	○	○	○	○
	The Milk Factor	○	○	○	●
	Unintended Relationships	○	○	○	○
	We Never Forget	○	○	○	○
Bösch et al. [5]	Address Book Leeching	●	●	●	●
	Bad Defaults	●	●	●	●
	Forced Registration	○	○	○	○
	Hidden Legalese Stipulations	●	●	●	●
	Immortal Accounts	○	○	○	○
	Information Milking	●	○	○	○
	Privacy Zuckering	●	○	●	○
	Shadow User Profiles	○	○	○	○
Gray et al. [19]	Forced Action	●	●	●	●
	Gamification	●	●	●	●
	Social Pyramid	●	●	●	●
	Interface Interference	●	●	●	●
	Aesthetic Manipulation	●	●	●	●
	False Hierarchy	●	●	●	●
	Hidden Information	●	●	●	●
	Preselection	●	●	●	●
	Toying With Emotions	●	●	●	●
	Nagging	●	●	●	●
	Obstruction	●	●	●	●
	Intermediate Currency	●	●	●	●
	Sneaking	●	●	○	○
Gray et al. [18]	Automating The User	●	●	○	●
	Controlling	●	●	●	●
	Entrapping	○	○	○	○
	Misrepresenting	●	●	●	●
	Nickling-And-Diming	○	○	○	○
	Two Faced	○	○	○	○
Mathur et al. [31]	Forced Acrtion (see Gray et al. [19])				
	Forced Enrollment	○	○	○	○
	Misdirection	●	●	●	●
	Pressured Selling	●	●	●	●
	Visual Interference	●	●	●	●
	Obstruction (see Gray et al. [19])				
	Hard To Cancel	●	●	●	●
	Scarcity	●	●	●	●
	High-Demand Messages	○	○	○	○
	Low-Stock Messages	○	○	○	○
	Sneaking (see Gray et al. [19])				
	Hidden Subscriptions	○	○	○	○
	Social Proof	●	●	●	●
	Activity Notifications	○	○	○	○
	Testimonials	○	○	○	○
	Urgency	●	●	●	●
	Countdown Timer	○	○	○	○
	Limited-Time Messages	○	○	○	○

Legend: **F** - Facebook
I - Instagram
Ti - TikTok
Tw - Twitter

Table 1: This table offers an overview of all 80 deductive codes used in the thematic analysis, sorted by authors. We only considered original sources for dark patterns to rely on unique codes and avoiding redundancy. As both Gray et al. [19] and Mathur et al. [31] include both high-level and low-level definitions, we used indentation to highlight categorical differences for these types of dark patterns. For transparency, we added three of Mathur et al.'s [31] high-level dark pattern categories, which were carried over from Gray et al. [19] and labeled accordingly. The codes were applied in four SNSs (F - Facebook, I - Instagram, Ti - TikTok, and Tw - Twitter). Codes that were identified by either of the two coders conducting the thematic analysis are indicated with "●" whereas "○" indicates that a code for a dark pattern was not found in the SNSs.

not yet covered. We decided early on to stick rigidly to established definitions of the types of dark patterns when applying deductive codes to avoid ambiguity among coders. During the coding of a first sample, we noticed issues were close in nature to certain deductive codes, but narrow wording hindered precise usage in different contexts. For instance, this was the case with the *confirmshaming* code. The underlying dark pattern, first coined by Brignull [7], describes texts that guilt users into certain actions based on shameful language. Applying an inverse strategy, we noticed various cases in which language was used in the form of positive encouragement to steer users towards a certain direction. We, therefore, created an additional code, *persuasive langugage*, that we defined bidirectionally describing any instance where language is used to push decisions. In total, we defined 22 unique codes describing problematic interactions or otherwise questionable interfaces that could not be coded using deductive codes (the entire inductive codebook, including code descriptions, are provided in Appendix A). As seen in Table 2, the 22 inductive codes did not occur across each of the four SNSs equally. Yet, 16 codes were applied to recording sessions on Facebook, 15 to TikTok and Twitter, while 14 were applied on Instagram sessions, showing a similar propagation of dark patterns across SNSs. Once all twelve sessions were analysed, we applied axial coding to the 22 codes. This resulted in the identification of five themes that describe the various types of dark patterns that are specific to SNSs. Answering our second research question, these themes encompass: (1) *interactive hooks*; (2) *social brokering*; (3) *decision uncertainty*; (4) *labyrinthine navigation*; and (5) *redirective conditions*. Moreover, we were able to assign these five themes to two broader strategies, to describe practitioners' and SNSs' intentions to navigate users' decision-making: (1) Engaging strategies and (2) Governing strategies. Engaging strategies envelope the themes *interactive hoooks* and *social-broking*, whilst governing strategies incorporate *decision uncertainty*, *labyrinthine navigation*, and *redirective conditions*. Table 2 provides a complete summary of these findings, including the 22 inductive codes, resulting in five themes and overarching strategies.

4.4 Engaging Strategies

In the context of SNSs, engaging strategies cover dark patterns where the goal is to keep users occupied and entertained for as long as possible. Prior dark patterns described by Roffarello and Russis [35] fall under this strategy alongside Lukoff et al.'s [28] discussion of potentially existing *attention-capture* dark patterns, which the OECD recently integrated [38]. As Table 2 indicates, the overall count of engaging strategies and subordinated codes is higher compared to those assigned to governing strategies, which is in line with research on people's motivation to use SNSs [46]. Emerging from the five themes, we identified two SNS-specific dark patterns - *interactive hooks* and *social brokering* - that subscribe to the engaging strategies (see Table 2).

4.4.1 Interactive Hooks. We define *interactive hooks* as design mechanisms that use rewarding schemes to keep users entertained and spend more time on a service. Throughout our recordings, we found multiple cases where such mechanisms were implemented. For example, some form of *gamification* elements [42] were coded within each SNS, galvanising users to share more information about themselves or connect to new people (see Figure 2). In another example, we found that many artefacts utilise *addictive mechanisms* [36] that often provide seemingly infinite content, further coded with *pull-to-refresh*, *infinite scrolling*, and *auto playing* media. With this finding, we confirm previous results of Lukoff et al. [29] who looked at auto-playing mechanisms on YouTube. We extend their work by also considering four alternative SNSs. By relying on a sequence of interactions, these addictive mechanisms can only be described with difficulty from still images, demonstrating the complexity in which dark patterns manifest in SNSs.

4.4.2 Social Brokering. We define *social brokering* as design mechanisms that nudge users to create multiple connections with people (e.g. based on similar characteristics) while suggesting new people to connect to, leading users to share more than they may want to a wider public. The name is inspired by agents whose aim is to facilitate connections between potential (transaction-) partners - here in the context of social networks. Although gamification strategies of the *interactive hooks* pattern already encourage users to increase their social connections, we found a range of artefacts specifically designed for this purpose codes as *social connector*. Moreover, we noticed that each SNS customised content presented based on the reviewers' usage behaviour. While in some instances the news feed content appeared a poor fit for the reviewers' preferences (coded as *false content customisation*), we also found very popular content reappearing across the recordings, captured in the code *regression toward the mean*. Visualised in Figure 3, Brignull's bad defaults pattern [7] is exploited to promote one's account to other users also outside the platform itself. Figure 4 shows how Instagram nudges its users to upload their device's local contacts to connect to more accounts quickly. However, by giving away details from private contacts, SNS providers can also store information from people who do not use their services without getting their consent first. In a similar case, Facebook uses persuasive language by telling users to add friends in order to see more content (see Figure 5).

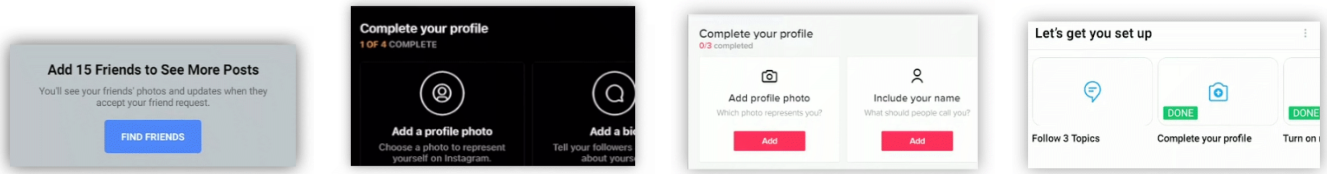
4.5 Governing Strategies

Governing strategies describe interface designs that navigate users' decision-making towards the designers' and/or platform providers' goals. Essentially, these are designed to control or govern user behaviour. While existing dark patterns, such as *interface interference* [17], fit this strategy's scope, we shine a light on not yet discussed dark pattern types: (1) *decision uncertainty*; (2) *labyrinthine navigation*; (3) and *redirective conditions*. All these strategies share a limitation of users' agency as SNS providers override users' goals with their own incentives.

4.5.1 Decision Uncertainty. We define *decision uncertainty* as strategies that are confusing to users by diminishing their ability to assess situations, leaving the user clueless as to what is expected of them or what options are available. Most similar to this dark pattern is the *confusing* dark pattern by Conti and Sobiesk [9]. However, in SNS, the strategy is not necessarily limited to incomprehensible questions or information but includes other elements, like *distraction* [9], that overwhelm users. Interface elements of the *decision uncertainty* code were so striking that we decided to promote it to a theme.

Strategy	Theme	No.	Inductive Code	F	I	Ti	Tw
Engaging Strategies	Interactive Hook	1.	Addictive Design	○	●	●	●
		2.	Autoplay Content	●	●	●	●
		3.	Fear Of Missing Out	○	○	○	●
		4.	Gamification	●	●	●	●
		5.	Infinite Scrolling	●	●	●	●
		6.	Pull To Refresh	●	●	●	●
		7.	Reduced Friction	●	●	●	●
	Social Brokering	8.	False Content Customisation	●	○	○	○
		9.	Regression Toward The Mean	●	○	●	●
		10.	Social Connector	●	●	●	○
Governing Strategies	Decision Uncertainty	11.	Decision Uncertainty	●	○	○	●
		12.	Clinging To Accounts	●	●	●	●
		13.	Persuasive Language	●	●	●	●
	Labyrinthine Navigation	14.	External Solution Search	○	●	○	○
		15.	Labyrinth	●	●	○	●
		16.	Hidden In Plain Sight	●	●	●	●
	Redirective Condition	17.	Auto Accept Third Party Terms	○	○	○	●
		18.	Decision Governing	●	●	○	○
		19.	Forced Access Granting	○	○	●	○
		20.	Forced Dialogue Interaction	●	○	●	○
		21.	Forced Grace Period	●	●	●	●
		22.	Plain Evil	○	○	●	○
			Total	16	14	15	15

Table 2: This table lists inductive codes and their presence within the four SNSs Facebook (F), Instagram (I), TikTok (Ti), and Twitter (Tw). Inductive codes observed by either of the two coders in a particular SNS session are visualised with "●" while "○" implies that a code was not found. This table further shows the themes each inductive code was assigned to based on axial coding of the data and the high-level strategies they subscribe to - engaging strategies and governing strategies. To align with prior research of this field, the developed themes are later referred to as types of dark patterns.



(a) Screenshot from Facebook **(b) Screenshot from Instagram** **(c) Screenshot from TikTok** **(d) Screenshot from Twitter**

Figure 2: Four examples of *interactive hooks*. Each sub-figure contains gamification elements that galvanise users to widen their social networks or publish more information about themselves.

Other elements obscured decision-making further. Some interfaces obfuscated the account deletion process (coded with *clinging to accounts*) while others used *persuasive language* to confuse users, similar to the *confirmshaming* dark pattern [7]. We found a quite unique design choice TikTok users faced when first logging on to the platform: When first opening the app after logging in, users are prompted with an interface asking them to choose preferred ad-related settings. While making their decision, both video and audio media is running in the background, contributing to cognitive overload. During the recordings, we noticed that reviewers of the data collection quickly complied with the platform's preference that utilises interface interference [17] and visual interference [31] dark patterns (see Figure 6). In a second example, Twitter users trying to delete their accounts will only find an option to deactivate it. Although it is possible to fully delete their account by following this path, the wording obfuscates this possibility.

4.5.2 Labyrinthine Navigation. We define *labyrinthine navigation* as nested interfaces that are easy to get lost in, disabling users from choosing preferred settings. This pattern is often seen in SNS settings menus. Related to *manipulating navigation* [9], but not necessarily steering users towards a designer's objective. Instead, this dark pattern describes interface architectures, such as menus, that users will easily get lost in, leaving them unsuccessful in achieving their goals. While recording the data, especially tasks six (visit the personal settings) and seven (visit the ad-related settings), surfaced

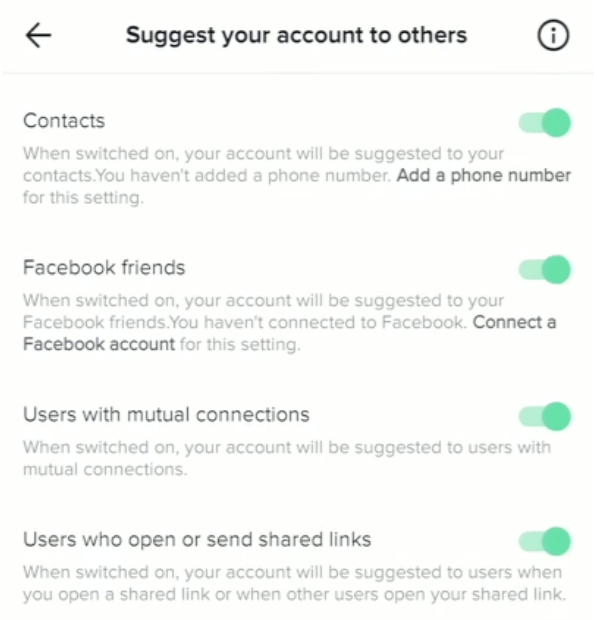


Figure 3: Example of *social brokering* from TikTok where settings are pre-set to suggest an account within and outside the SNS. The interface includes *bad defaults* and *privacy zuckering* dark patterns

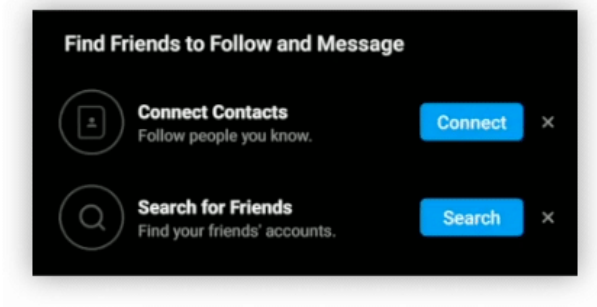


Figure 4: Example of *social brokering* from TikTok where settings are pre-set to suggest an account within and outside the SNS.

difficulties for reviewers to find specific settings, some of them coded *hidden in plain sight* camouflaged between a wide collection of other options. We noticed this issue across all four SNSs. In a particular case, we noticed one participant using Instagram to switch applications to an online search engine to look up the solution for how to find a specific setting.

4.5.3 Redirective Conditions. We define *redirective condition* as choice limitations that force users to overcome unnecessary obstacles before being able to achieve their goals. Redirective conditions usually favour the objectives of the SNS. The *forced action* dark pattern [17] plays an important role here, as users are required to first comply with certain demands before being able to do what they want. In the context of SNS, this dark pattern includes passive functionalities, like the restriction of services that are only lifted

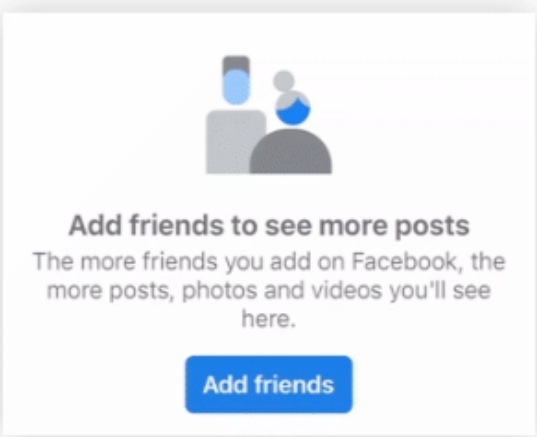


Figure 5: Example of *social brokering* from Facebook. Once users reach the bottom of their timeline content, this message appears, nudging them to add friends to see new posts.

after users give permissions unrelated to the functionality (coded as *forced access granting*). In related situations, we noticed *forced dialogue interactions* that required users to engage with text elements otherwise occupying desired functionalities. More specific yet similar to the *roach motel* [7], *hard to cancel* [31], or *account deletion roadblocks* [21] dark patterns, a grace period of up to 30 days was announced by all SNSs before certain deletion processes were accepted. The *Forced Grace Period* code was used in such instances. When deleting content, Facebook keeps the targeted item, as seen in Figure 7. Trying to bypass this rule, participants found it difficult to find the particular settings that would delete their content immediately. Interestingly, all four SNS denied users an immediate account deletion, as demonstrated by TikTok in Figure 8. Each had a 30 days return option that would automatically reactivate accounts, even if users accidentally logged in to the SNS.

5 DISCUSSION & IMPLICATIONS

Our examination of four SNSs — Facebook, Instagram, TikTok, and Twitter — identified instances of 44 dark patterns from a taxonomy of 80 codified in prior dark pattern literature. The work also identified instances of 22 inductive codes capturing malicious design artefacts that were not outlined in earlier research. Findings stem from expert reviews of the aforementioned SNSs carried out by six trained HCI researchers. Each researcher reviewed two SNSs and executed ten tasks designed to thoroughly understand the applications. The 22 inductive codes were then analysed, supported by axial coding. This resulted in the generation of five common themes describing SNS-specific dark patterns and two high-level strategies: Engaging strategies and governing strategies. Figure 9 summarises these findings while further allocating Mathur et al.'s [32] dark pattern characteristics.

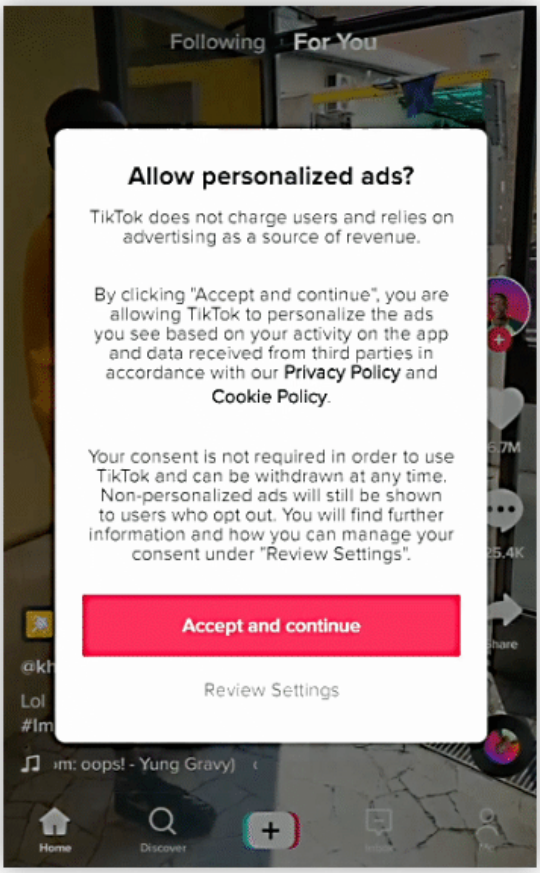


Figure 6: Example of *decision uncertainty* from TikTok. After first logging in, users of the SNS are asked to choose preferences for their ad-related settings. Meanwhile, video and audio media are playing in the background, complicating the interaction. The interface further contains *interface interference, visual interference* and hidden-legalese stipulation dark patterns.

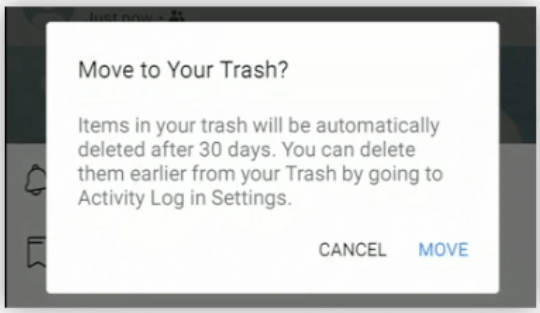


Figure 7: Example of *redirective conditions* from Facebook. When trying to delete a post, Facebook will instead keep the item stored for another 30 days.

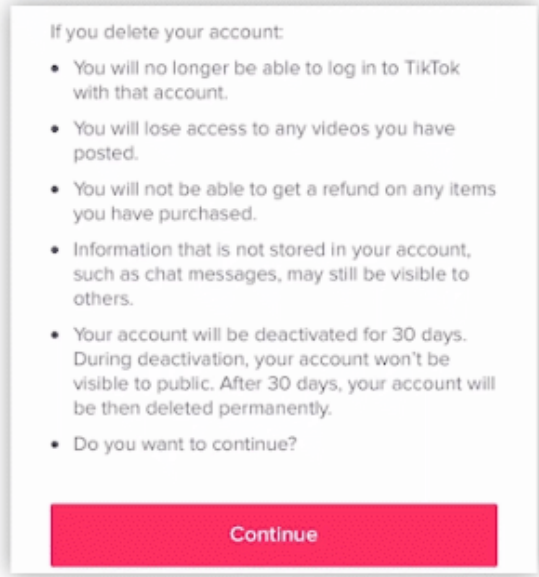


Figure 8: Example of *redirective conditions* from TikTok. Accounts will be kept and able to be restored for 30 days without giving users the choice to bypass this decision.

5.1 Dark Patterns In SNSs

To explore the potential existence of dark patterns in SNSs, we began the thematic analysis by initialising a deductive dark pattern codebook comprising 80 dark pattern types from prior works [5, 7, 9, 18–20, 31, 47]. Answering our first research question regarding the types of dark patterns that are used across the four SNSs, we identified instances of 44 types. This might suggest that the other 36 types of dark patterns are, as our experience conducting thematic analysis attests, highly domain-specific, hindering their potential to be applied elsewhere.

Indeed, despite the relative success of our attempts to apply these dark patterns to SNSs, different levels of generalisability and specification caused ambiguity decreasing confidence of coders when applying certain dark patterns. Some works provide alternating abstractions for their taxonomies. Conti and Sobiesk [9], for instance, call their patterns malicious interface design *techniques*, as the term dark pattern was not widely established at the time of their publication. In another work, Gray et al. [19] speak of dark pattern *strategies*, placing their findings on a more abstract level as, for example, Zagal et al.' [47] dark patterns, who speak of game dark patterns without further labels.

Because we agree with the necessity to understand dark patterns in different hierarchical contexts, we chose to include all taxonomies in our study to learn about differences when they are applied in situations outside their original field of application. Further guided by our decision to remain close to provided definitions,

we learned that dark patterns with abstract definitions were applied more often throughout the SNSs. On the contrary, 36 dark patterns were left unused during our study. Certain dark patterns shared a name with varying definitions leading to ambiguity when used. Some dark patterns were derived from highly specific contexts, such as e-commerce [7, 31] or games [47], which hindered their usage elsewhere. Others contained overly precise descriptions descriptions [5] but could find usage elsewhere if phrased more generically. This implies that general and unspecific dark patterns gain the utility to describe unethical practices and, thus, better help to identify interface problems.

This implication is in line with prior research by Mathur et al. [31]. Aside from the identification of 12 dark patterns, the authors found an alternative approach to abstract the basic operations of dark patterns by mapping them onto cognitive biases that they exploit [31, 32]. The authors developed six characteristics distinguishing overall approaches to dark patterns: (1) asymmetric; (2) covert; (3) deceptive; (4) hides information; (5) restrictive; and (6) disparate treatment. Through these characteristics, the authors were able to characterise a comprehensive dark pattern taxonomy in which we notice an accessible and extendable framework for future works.

5.2 SNS-Specific Dark Patterns & Strategies

Answering the second research question, asking about dark patterns unique to SNSs, our study revealed five types of dark patterns not contained in previous work. When defining the dark patterns, one goal was to provide enough abstraction to enable applicability outside SNSs. For the same reason, we placed them in Mathur et al.'s [31, 32] six characteristics to allow future works to consider these dark patterns under their lens as well. As can be seen in Figure 9, not all of Mathur et al.'s characteristics could be applied with the same effectiveness or explicitness. Interestingly, we were not able to apply the *disparate treatment* characteristic at all because it solely contains Zagal et al.'s [47] game dark patterns which describe mechanics where games benefit players who, for example, buy advantages. We found none of the four considered SNSs to offer distinct benefits to certain users while posing disadvantages to others.

Another reason why Mathur et al.'s characteristics show lower effectiveness in SNSs may be described in how the two strategies operate, distinguishing them from prior dark pattern settings. As an incentive to increase time spent on their platforms, SNS need satisfied users. Since the implementation of harmful designs could act as an antagonist to this goal, jeopardising potential advertisement revenue, we describe engaging and governing strategies that navigate users' decision-making while possibly keeping their satisfaction with the SNS high [34]. Although we cannot draw an easy causal connection between how dark patterns affect users and users' well-being, assessing identified SNS dark patterns further resulted in the description of two design strategies, *engaging strategies* and *governing strategies* (see Figure 9). While one group aims to increase the time users spend on SNS, the other governs their decision-making by nudging users into desired directions or keeping them from other options by, for example, designing for increased friction.

With the intent to enable future work to better build on our findings and inspired by the types of dark patterns that showed more pertinence in our study, our aim was to describe our dark patterns to allow application outside SNS-specific contexts. As a different environment, many games feature achievement tracks requiring players to regularly play the game to progress. Similar to Zagal et al.'s *playing by appointment*, it is imaginable that these games feature *interactive hooks* to keep players engaged or contain distracting elements obfuscating users' decisions and, thus, deploying *decision uncertainty*.

Although this work mainly contributes to the academic community, this strand of research has gained traction outside, including regulation and guidelines. In a first line of defence against violations of GDPR [8] requirements in SNS contexts, the EDPB [3] developed a guideline for designers and users to recognise and avoid dark patterns. However, the included dark pattern categories (overloading, skipping, stirring, hindering, fickle, and left in the dark) lack alignment with taxonomies developed in the academic community that relies on empirical evidence. In this regard, we acknowledge a potential for better cooperation between law and HCI research to work on the protection of users in a joint effort. To offer some aid in these efforts, our goal was to create definitions that are broadly applicable. The five dark patterns expand current taxonomies by the scope of SNSs, whereas the two strategies provide high-level categories to describe alternate burdens placed on users not yet covered.

6 LIMITATIONS & FUTURE WORK

As previously stated in Section 3, we faced certain limitations in the data collection phase of this work. Firstly, we decided to focus on only four SNSs that we deemed comparable after prior considerations looking at a wider range of applications for the user count and similar functionalities. This limitation extends to the fact that we limited the review to only looking at their mobile applications. Although this limitation allowed us to study each SNS in a single modality on a deeper level, our findings do not necessarily represent all SNSs. None of the considered SNSs indicates any usage of the *disparate treatment* characteristic, which describes discrepancies in the treatment of paid and free customers, respectively. This might be more relevant in a study of LinkedIn, for example, as it offers both free and paid subscriptions, the latter of which affords certain advantages like hiding one's identity when looking at another user's profile. Future work could extend our findings by considering other SNSs. Secondly, we decided early on to only focus on mobile applications of considered SNSs, following users' preferences. However, users may be faced with alternative dark patterns based on SNSs' modalities, as the research by Guanawa et al. [21] or Schaffner et al. [39] suggest. Future studies could consider our results when expanding on other SNSs and modalities. Thirdly, this research was conducted during the COVID-19 pandemic. The data collection was thus conducted without supervision after providing necessary information, including a comprehensive manual. While neither screen nor voice recordings indicated confusion about the tasks, room for misunderstanding remained. Thirdly, we relied on HCI researchers as expert reviewers. This decision is

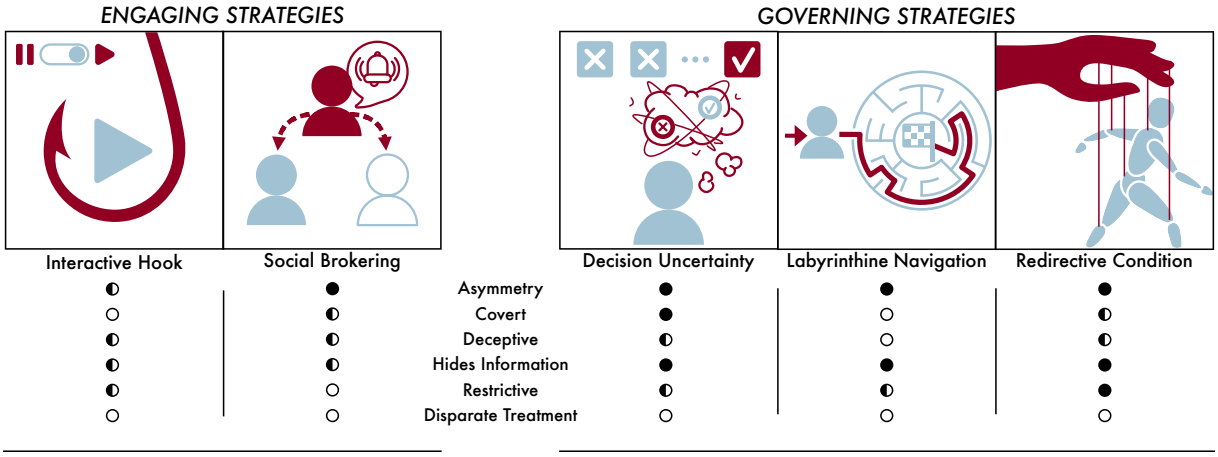


Figure 9: Summary of engaging and governing strategies with five SNS dark patterns in two strategies - engaging and governing. For future work, each dark pattern was assigned corresponding attributes following Mathur et al. [31] six dark pattern characteristics.

justified by their competence to understand and recognise state-of-the-art design practices, unlike regular users would be able to. However, by limiting reviewers' expertise to a single domain, we may have missed alternative expertise that may surface additional findings. Future work could consider recruiting experts with backgrounds in cognitive science and psychology, including a thorough understanding of cognitive biases. Their expertise could be particularly interesting in establishing connections between our current understanding of dark patterns and of cognitive biases.

During the thematic analysis, we also faced some limitations. Firstly, we limited the deductive codebook to eight works comprising 81 dark patterns. We relied on Mathur [32] dark pattern review to get a comprehensive taxonomy. However, we decided to neglect dark patterns that were described outside the academic community, including guidelines such as those published by the NCC [10], CNIL [14], or the EDPB [3]. Moreover, dark patterns by Gunawan [21] were not included, as they were not mapped onto Mathur et al. [31, 32] dark pattern characteristics, making comparisons more difficult. Future work could generate a complete mapping of a comprehensive dark pattern taxonomy onto Mathur et al.'s characteristics. Connecting dark patterns to cognitive biases, as Mathur et al. propose, seems to be a natural next step for the dark pattern discourse. To prepare this work for this direction, we assigned our findings to appropriate characteristics. In this process, we noticed that the current scope of characteristics does not cover all problematic design strategies contained in SNS. However, a thorough analysis of potential cognitive biases active in SNS would have been outside the scope of this research. Future work could address this gap and extend this work by bringing together the studies surrounding cognitive biases and dark patterns to generate a resourceful and sustainable framework.

7 CONCLUSION

In recent years, the dark pattern landscape has expanded into various different domains. In this paper, we contribute to this body of research through the application and expansion of a comprehensive taxonomy of dark patterns in the context of SNSs. Supported by thematic analysis, we investigate Facebook, Instagram, TikTok, and Twitter and confirm that the platforms all deployed a variety of dark patterns and design strategies aimed at limiting users' agency, steering their decision-making. Findings suggest that dark patterns with higher grades of abstraction are easier to be applied in multiple contexts compared to those given narrower definitions. Lastly, our results yield evidence for two high-level strategies - engaging and governing - containing five types of dark patterns previously not described.

ACKNOWLEDGMENTS

The research of this work was partially supported by the Klaus Tschira Stiftung gGmbH.

REFERENCES

[1] Ine Beyens, J Loes Pouwels, Irene I van Driel, Loes Keijsers, and Patti M Valkenburg. 2020. The effect of social media on well-being differs from adolescent to adolescent. *Scientific Reports* 10, 1 (2020), 1–11.

[2] Ann Blandford, Dominic Furniss, and Stephann Makri. 2016. Qualitative HCI Research: Going Behind the Scenes. *Synthesis Lectures on Human-Centered Informatics* 9, 1 (April 2016), 1–115. https://doi.org/10.2200/S00706ED1V01Y201602HCI034 Publisher: Morgan & Claypool Publishers.

[3] European Data Protection Board. March, 2022. Guidelines 3/2022 on Dark patterns in social media platform interfaces: How to recognise and avoid them | European Data Protection Board. https://edpb.europa.eu/our-work-tools/documents/public-consultations/2022/guidelines-32022-dark-patterns-social-media_en Visited on 2022-03-29.

[4] Kerstin Bongard-Blanchy, Arianna Rossi, Salvador Rivas, Sophie Doublet, Vincent Koenig, and Gabriele Lenzini. 2021. "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* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 763–776. https://doi.org/10.1145/3461778.3462086

[5] Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher. 2016. Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns. *Proc. Priv. Enhancing Technol.* 2016, 4 (2016), 237–254.

[6] Virginia Braun, Victoria Clarke, Nikki Hayfield, and Gareth Terry. 2019. Thematic Analysis. In *Handbook of Research Methods in Health Social Sciences*, Pranee Liamputtong (Ed.). Springer Singapore, Singapore. https://doi.org/10.1007/978-981-10-5251-4_103

[7] Harry Brignull. 2010. Deceptive Design – formerly darkpatterns.org. https://www.deceptive.design/ Visited on 2022-03-29.

[8] European Commission. 2016. GDPR-16 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement

of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-changes_en.pdf

[9] Gregory Conti and Edward Sobiesk. 2010. Malicious interface design: exploiting the user. In *Proceedings of the 19th international conference on World wide web - WWW '10*. ACM Press, Raleigh, North Carolina, USA, 271. https://doi.org/10.1145/1772690.1772719

[10] Frobrukerrådet (Norwegian Consumer Council). 2019. Deceived by design: How tech companies use dark patterns to discourage us from exercising our rights to privacy. https://fil.forbrukerradet.no/wp-content/uploads/2018/06/2018-06-27-deceived-by-design-final.pdf5

[11] Ratan Dey, Zubin Jelveh, and Keith Ross. 2012. Facebook users have become much more private: A large-scale study, In 2012 IEEE International Conference on Pervasive Computing and Communications Workshops, PERCOM Workshops 2012. *2012 IEEE International Conference on Pervasive Computing and Communications Workshops, PERCOM Workshops 2012* 01, 346–352. https://doi.org/10.1109/PerComW.2012.6197508

[12] Linda Di Geronimo, Larissa Braz, Enrico Fregnan, Fabio Palomba, and Alberto Bacchelli. 2020. UI Dark Patterns and Where to Find Them: A Study on Mobile Applications and User Perception. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376600

[13] Sindhu Kiranmai Ernala, Moira Burke, Alex Leavitt, and Nicole B. Ellison. 2020. How Well Do People Report Time Spent on Facebook? An Evaluation of Established Survey Questions with Recommendations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376435

[14] Commission Nationale Informatique et Libertés. 2019. On the practical procedures for collecting the consent provided for in article 82 of the french data protection act, concerning operations of storing or gaining access to information in the terminal equipment of a user (recommendation "cookies and other trackers"). https://www.cnil.fr/sites/default/files/atoms/files/draft_recommendation_cookies_and_other_trackers_en.pdf

[15] Lothar Fritsch. 2017. Privacy dark patterns in identity management. In *Open Identity Summit 2017 : Proceedings (Lecture Notes in Informatics, 277)*. Gesellschaft für Informatik, Bonn, Germany, 93–104.

[16] ATLAS.ti Scientific Software Development GmbH. 2021. ATLAS.ti: The Qualitative Data Analysis & Research Software. https://atlasti.com/

[17] Colin M. Gray and Shruthi Sai Chivukula. 2019. Ethical Mediation in UX Practice. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3290605.3300408

[18] Colin M. Gray, Shruthi Sai Chivukula, and Ahreum Lee. 2020. *What Kind of Work Do "Asshole Designers" Create? Describing Properties of Ethical Concern on Reddit*. Association for Computing Machinery, New York, NY, USA, 61–73. https://doi.org/10.1145/3357236.3395486

[19] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. *The Dark (Patterns) Side of UX Design*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3174108

[20] Saul Greenberg, Sebastian Boring, Jo Vermeulen, and Jakub Dostal. 2014. *Dark Patterns in Proxemic Interactions: A Critical Perspective*. Association for Computing Machinery, New York, NY, USA, 523–532. https://doi.org/10.1145/2598510.2598541

[21] Johanna Gunawan, Amogh Pradeep, David Choffnes, Woodrow Hartzog, and Christo Wilson. 2021. A Comparative Study of Dark Patterns Across Web and Mobile Modalities. *Proceedings of the ACM on Human-Computer Interaction* 5 (2021), 1–29. Issue CSCW2. https://doi.org/10.1145/3479521

[22] Hana Habib, Sarah Pearman, Ellie Young, Ishika Saxena, Robert Zhang, and Lorrie FaIth Cranor. 2022. Identifying User Needs for Advertising Controls on Facebook. *Proceedings of the ACM on Human-Computer Interaction* 6 (2022), 1–42. Issue CSCW1. https://doi.org/10.1145/3512906

[23] Edward T. Hall. 1966. *The hidden dimension*. Doubleday, Garden City, N.Y.

[24] Peter Hustinx. 2010. Privacy by design: delivering the promises. *Identity in the Information Society* 3, 2 (Aug. 2010), 253–255. https://doi.org/10.1007/s12394-010-0061-z

[25] Monique W.M. Jaspers, Thiemo Steen, Cor van den Bos, and Maud Geenen. 2004. The think aloud method: a guide to user interface design. *International Journal of Medical Informatics* 73, 11 (2004), 781–795. https://doi.org/10.1016/j.ijmedinf.2004.08.003

[26] Reynol Junco. 2013. Comparing actual and self-reported measures of Facebook use. *Computers in Human Behavior* 29, 3 (May 2013), 626–631. https://doi.org/10.1016/j.chb.2012.11.007

[27] California State Legislature. 2018. CCPA-18 2018. California Consumer Privacy Act of 2018 [1798.100 - 1798.199] (CCPA). https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5.

[28] Kai Lukoff, Cissy Yu, Julie Kientz, and Alexis Hiniker. 2018. What Makes Smartphone Use Meaningful or Meaningless? *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 22 (March 2018), 26 pages. https://doi.org/10.1145/3191754

[29] Kai Lukoff, Cissy Yu, Julie Kientz, and Alexis Hiniker. 2018. What Makes Smartphone Use Meaningful or Meaningless?, In Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 22, 26 pages. https://doi.org/10.1145/3191754

[30] Maximilian Maier. 2020. Dark Design Patterns - An End-user Perspective. *Human Technology* 16 (2020), 170–199. https://doi.org/10.17011/ht/urn.202008245641

[31] Arunesh Mathur, Gunes Acar, Michael J. Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark Patterns at Scale. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (nov 2019), 1–32. https://doi.org/10.1145/3359183

[32] Arunesh Mathur, Mihir Kshirsagar, and Jonathan Mayer. 2021. *What Makes a Dark Pattern... Dark? Design Attributes, Normative Considerations, and Measurement Methods*. Association for Computing Machinery, New York, NY, USA, pp. 1–18. https://doi.org/10.1145/3411764.3445610

[33] Philipp Mayring. 2020. Qualitative Inhaltsanalyse. In *Handbuch Qualitative Forschung in der Psychologie: Band 2: Designs und Verfahren*, Günter Mey and Katja Mruck (Eds.). Springer Fachmedien Wiesbaden, Wiesbaden, 495–511. https://doi.org/10.1007/978-3-658-26887-9_52

[34] Thomas Mildner and Gian-Luca Savino. 2021. Ethical User Interfaces: Exploring the Effects of Dark Patterns on Facebook. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3411763.3451659

[35] Alberto Monge Roffarello and Luigi De Russis. 2022. Towards Understanding the Dark Patterns That Steal Our Attention. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. ACM, New Orleans LA USA, 1–7. https://doi.org/10.1145/3491101.3519829

[36] Christian Montag, Bernd Lachmann, Marc Herrlich, and Katharina Zweig. 2019. 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 (Jan. 2019), 2612. https://doi.org/10.3390/ijerph16142612 Number: 14 Publisher: Multidisciplinary Digital Publishing Institute.

[37] Jakob Nielsen. 1994. Usability inspection methods. In *Conference Companion on Human Factors in Computing Systems (CHI '94)*. Association for Computing Machinery, New York, NY, USA, 413–414. https://doi.org/10.1145/259963.260531

[38] OECD. 2022. Dark commercial patterns. , 96 pages.

[39] Brennan Schaffner, Neha A. Lingareddy, and Marshini Chetty. 2022. Understanding Account Deletion and Relevant Dark Patterns on Social Media. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 1–43. https://doi.org/10.1145/3555142

[40] Sarita Yardi Schoenebeck. 2014. Giving up Twitter for Lent: how and why we take breaks from social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 773–782. https://doi.org/10.1145/2556288.2556983

[41] Holly B. Shakya and Nicholas A. Christakis. 2017. Association of Facebook Use With Compromised Well-Being: A Longitudinal Study. *American Journal of Epidemiology* 185, 3 (Feb. 2017), 203–211. https://doi.org/10.1093/aje/kww189

[42] Jorge Simões, Rebeca Díaz Redondo, and Ana Fernández Vilas. 2013. A social gamification framework for a K-6 learning platform. *Computers in Human Behavior* 29, 2 (March 2013), 345–353. https://doi.org/10.1016/j.chb.2012.06.007

[43] Jin-Liang Wang, Linda A. Jackson, James Gaskin, and Hai-Zhen Wang. 2014. The effects of Social Networking Site (SNS) use on college students' friendship and well-being. *Computers in Human Behavior* 37 (Aug. 2014), 229–236. https://doi.org/10.1016/j.chb.2014.04.051

[44] Yang Wang, Pedro Giovanni Leon, Kevin Scott, Xiaoxuan Chen, Alessandro Acquisti, and Lorrie Faith Cranor. 2013. Privacy nudges for social media: An exploratory facebook study. *WWW 2013 Companion - Proceedings of the 22nd International Conference on World Wide Web* 01 (2013), 763–770.

[45] Yang Wang, Gregory Norcie, Saranga Komanduri, Alessandro Acquisti, Pedro Giovanni Leon, and Lorrie Faith Cranor. 2011. "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* (Pittsburgh, Pennsylvania) *(SOUPS '11)*. Association for Computing Machinery, New York, NY, USA, Article 10, 16 pages. https://doi.org/10.1145/2078827.2078841

[46] L. Y.C. Wong and Jacquelyn Burkell. 2017. Motivations for Sharing News on Social Media. In *Proceedings of the 8th International Conference on Social Media & Society (#SMSociety17)*. Association for Computing Machinery, New York, NY, USA, 1–5. https://doi.org/10.1145/3097286.3097343

[47] José P Zagal, Staffan Björk, and Chris Lewis. 2013. Dark Patterns in the Design of Games. In *Proceedings of the 8th International Conference on the Foundations of Digital Games (FDG 2013)* (May 14-17). Society for the Advancement of the Science of Digital Games, Chania, Crete, Greece, 39–46. http://www.fdg2013.org/program/papers.html

A INDUCTIVE CODEBOOK INCLUDING DESCRIPTIONS

	Code	Description
1	Addictive Design	Features or elements that keep users hooked to content.
2	Auto Accept Third Party Terms	Unknowingly giving consent to share data with third parties per default settings.
3	Autoplay Content	Auto-playing content without further actions by the user.
4	Clinging To Accounts	Making the process of deleting an account unnecessarily difficult or reactivating accounts after deletion has already been initialised.
5	Decision Governing	Interface instances that navigate or steer users' decision-making.
6	Decision Uncertainty	Users do not know what they are left in confusion and what consequences their decision will have.
7	External Solution Search	Not able to find specific features or settings, users fall back to use search engines to find what they are looking for.
8	False Content Customisation	Shown content does not fit the users' preferences or followed accounts.
9	Fear Of Missing Out	Feeling pressured to (re)visit specific SNS features out of fear to miss something.
10	Forced Access Granting	Features requiring unnecessary access to special device hardware or local data.
11	Forced Dialogue Interaction	Prompting Text Boxes that require immediate attention with no option to dismiss them without interaction.
12	Forced Grace Period	When deleting content or accounts, users are forced to wait a minimum amount of days before changes become active (often 30 days).
13	Gamification	Playful elements that motivate users to do something by presenting artificial progress and/or rewards to get more data.
14	Hidden In Plain Sight	Crucial information is often obscured by attention-grabbing interface elements.
15	Infinite Scrolling	Users can infinitely scroll through content (often, more and more suggested content is shown the more they progress.
16	Labyrinth	Nested interface structures users get easily lost in.
17	Persuasive Language	Emotional pressuring language to push towards a specific direction that may not be in the users best interest.
18	Plain Evil	Interface elements suddenly change functionalities or are being exchanged for alternative ones.
19	Pull To Refresh	Pulling downward on a content-displaying interfaces will load new content. Sometimes, suggested content is shuffled into the feed to give users more to look at.
20	Reduced Friction	Purposefully making certain elements easier accessible and thus pushing alternatives into the back.
21	Regression Toward The Mean	Users are suggested topics which most people like and thus are likely to add to the already existing "popular main topics".
22	Social Connector	Requesting additional information to connect to friends/family and other social circles.

Table 3: Table containing all 22 codes from the inductive codebook from study 1, later used to create five themes deriving into the SNS-specific dark patterns.