

Hell is Paved with Good Intentions: The Intricate Relationship Between Cognitive Biases and Dark Patterns

THOMAS MILDNER, University of Bremen, Germany
 ALBERT INKOOM, University of Bremen, Germany
 RAINER MALAKA, University of Bremen, Germany
 JASMIN NIESS, University of Oslo, Norway

Throughout the past decade, research in HCI has identified numerous instances of dark patterns in digital interfaces. These efforts have led to a well-fostered typology describing harmful strategies users struggle to navigate [33]. However, an in-depth understanding of the underlying mechanisms that deceive, coerce, or manipulate users is missing. We explore the interplay between cognitive biases and dark patterns to address this gap. To that end, we conducted four focus groups with experts ($N = 15$) in psychology and dark pattern scholarship, inquiring how they conceptualise the relation between cognitive biases and dark patterns. Based on our results, we constructed the “Relationship Model of Cognitive Biases and Dark Patterns” which illustrates how cognitive bias and deceptive design patterns relate and identifies opportune moments for ethical reconsideration and user protection mechanisms. Our insights contribute to the current discourse by emphasising ethical design decisions and their implications in the field of HCI.

Additional Key Words and Phrases: cognitive bias, deceptive design patterns, dark patterns, design ethics, responsible design

ACM Reference Format:

Thomas Mildner, Albert Inkoom, Rainer Malaka, and Jasmin Niess. 2024. Hell is Paved with Good Intentions: The Intricate Relationship Between Cognitive Biases and Dark Patterns. In . ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

Draft: May 12, 2024

1 INTRODUCTION

The last decade of research in Human-Computer Interaction (HCI) has shown a growing interest in unethical design practices — in the midst of it are deceptive design practices and so-called “dark patterns”¹. Various streams of research have described instances that deceive or manipulate users in domains such as, but not limited to, e-commerce [58], social media [64, 71], and web and mobile interfaces [22, 32, 38]. While this body of work makes significant contributions that inform the protection of people against the harms embedded in these environments, we currently lack a fundamental understanding of the underlying principles that enable the deceptive, coercive, and potentially manipulative characteristics of dark

¹We opted for the term “dark pattern” in alignment with previous authors when referring to identified harmful design practices. However, we are aware that the ACM Diversity, Equity, and Inclusion Council recently decided to classify the term as problematic, given its association with negative connotations. We use the term in line with its original context of *hidden* consequences for users [11] and acknowledge that there currently exists no alternative terms that convey the full spectrum of deceptive, manipulative, obstructive, or coercive characteristics of the original term.

Authorversion, 2024.

© 2024 Copyright held by the owner/author(s).

This is the author’s version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in , <https://doi.org/10.1145/nnnnnnn.nnnnnnn>.

patterns. Previous work by Mathur et al. [58] and Waldman [82] draw tentative connections to the exploitation of cognitive biases in this regard. However, the lack of supportive research leaves a gap for additional, fundamental work to describe this relationship in more detail. Furthermore, it remains unclear how this relationship could manifest itself and how it could be navigated in interaction design.

Generally, exploiting cognitive biases and connected heuristics [5, 50] to manipulate users of any kind poses important ethical questions that require careful evaluation. Incorporating knowledge of human cognition and perception is an integral element of human factors and the design of interfaces. However, Thaler and Sunstein’s ‘nudge’ theory [76] and the principles of persuasive design [26] illustrate the effectiveness in which cognitive biases can be utilised to alter peoples’ choice architecture; and, thus, require responsibility in designers not to hinder informed decision-making as users can be unaware of consequences [16]. Wary of misuse, critique voices ethical concerns against paternalistic implications on agency [43, 60, 70]. Although Thaler and Sunstein independently addressed these concerns [73, 75] by reiterating how nudges were meant to empower people to make individual good decisions, Hansen and Jespersen [40] argued for the necessity of nudging transparently upholding user agency.

This ongoing discourse spotlights the various ethical caveats designers should consider to support the informed decision-making of end-users. While recent work shows positive effects of design friction to assist users in making informed decisions [53, 61], in practice, designers aim for effective user journeys that meet their goals. Online shopping sites, for example, take advantage of the aforementioned principles [16, 58]; they steer users toward effortlessly discovering (recommended) products from where they are quickly manoeuvred to a checkout page [58]. In other cases, design friction is indeed used to hinder certain interactions. This can be sensible to support reflection of an interaction’s consequences by interrupting mindless engagement [19, 61]. However, design friction can lead to frustration in other cases. Along similar lines, infamous cookie-consent banners have a history of making selection processes a cumbersome task [34]. They rely on visual interferences and aesthetic manipulation to pre-select choices, effectively pushing users’ decisions to their disadvantage [34].

These examples demonstrate the responsibility contained in designers’ work. Exploitation of cognitive biases, such as the framing effect or the default effect, have been used to misguide and trick users [58]. Drawing from the importance of this ongoing discourse and the need for a better understanding of the underlying cognitive mechanisms that can be exploited to harm users, this research aims

to take the next step by exploring the relationship between cognitive biases and dark patterns and how it should be navigated in interaction design. To this end, we conducted four focus groups with experts in dark pattern scholarship, cognitive science, and psychology research, several of whom also have significant experience in interaction design ($N = 15$). Each focus group included a discussion structured to investigate the following research question:

RQ: How can we conceptualise the dynamic relationship between cognitive biases and deceptive design patterns?

In conclusion, the focus groups facilitated valuable insights by bridging the perspectives of two distinct fields, providing important details to understand the effects of cognitive biases on dark patterns and vice versa. Moreover, our findings offer guidance for future work and ethical design considerations to better protect users. By examining the inherent characteristics of cognitive biases [78] and their utilisation in dark patterns within digital contexts, we have garnered new insights into underlying mechanisms, thereby supporting previous suggestions that linked the harming effects of dark patterns to the exploitation of cognitive biases [58, 82]. Our contribution to this field is the *Relationship Model of Cognitive Biases and Dark Patterns*, which elucidates the connections between design decisions and real-world consequences, emphasizing the necessity of responsible design choices and decision-making to protect end-users.

2 RELATED WORK

This paper synthesises research on autonomy and empowerment, cognitive biases, and dark patterns. First, we revisit contributions about users' autonomy within the periphery of HCI promoting ethical design concepts. Second, we present a background of cognitive bias scholarship that informed this research. Concluding this section, the third part focuses on HCI research on dark patterns and points to an ongoing discourse regarding terminology in this area.

2.1 User Autonomy and Empowerment

As digital technologies have become increasingly ubiquitous, the relationship people share with their daily drivers has been described as complex and emotional [74]. The field of HCI has advanced from focusing on usability issues to designing technologies that foster positive interactions and user well-being, often navigating persuasion and autonomy. As an antagonist to autonomy, persuasive technologies often undermine user agency [15]. While research in health-related environments (e.g. [4]) indicates potential positive applications of the persuasive technology paradigm, it is crucial to remain mindful of the ongoing critiques and ethical considerations surrounding persuasive technology. At the same time, it is important to recognise that participants made an autonomous decision and consented before giving up autonomy to change their behaviour toward more healthy options using this health-care intervention.

Unfortunately, achieving responsible design is not an easy task [37]. While Schneider et al. [72] emphasised the importance of empowerment based on a structured literature review, persuasive design remains an often relied-on strategy for behaviour change interventions [39]. Thereby, the implementation of persuasive design is often well-meant but executed in a problematic manner [15].

To illustrate, in their work, Brynjarsdóttir et al. [12] describe the pitfalls of persuasive sustainability — the attempt to persuade users' behaviour toward environmental sustainability, related to Fogg's Behaviour Model [26]. In sum, Brynjarsdóttir et al. frame a critic concerned with modernist technologies, which they place persuasive sustainability under, and remind about a lack of awareness of a design's impact. Opening future avenues, the authors suggest that, firstly, persuasion and users' reactions have to be better understood before implementation. Secondly, they echo traditional user-centred design (UCD) approaches by reminding practitioners to involve users throughout development stages, and, thirdly, encourage practitioners to move away from the individual and consider larger social environments. Offering both research and industry a reflection tool in this vein, Elsayed-Ali et al. [24] created an online card tool showcasing a variety of critical aspects and considerations for innovative design processes. To that end, the tool fosters the sharing of opinions across hierarchies, accounts for differences among participants, promotes inclusiveness, and gives room to otherwise difficult discussions specific to the participants' work environments.

Work in HCI has long been concerned with conscious control [48], and in that sense user autonomy and agency; although often using these terms interchangeably, as highlighted by a recent literature review [7] considering over 32 years of related work. In this review, Bennet et al. [7] allocate these values as key concepts in various HCI work. In 2014, Grimpe et al. [37] highlighted the difficulties of responsible design, problematizing the challenge into four core issues: *reflexivity*; *responsibility and responsiveness*; *inclusion*; and *anticipation*. A core effort promoting user autonomy when interacting with technologies is value sensitive design (VSD), first proposed by Friedman et al. [27]. Continuing former work [28], the authors encourage interfaces to convey functionalities transparently and truthfully while empowering users. The relevance of VSD is again highlighted by contemporary work from Chen et al. [16]. The authors apply underlying paradigms to often deceptive recommender systems to study notions of disagreement between practitioners and users, illuminating disagreement of associated values. In part, these disagreements may be the result of conflicting ethical stances among practitioners and corporate incentives [17, 31].

Resulting interfaces might restrict users' ability to make informed decisions that are in line with their beliefs or values [26, 84, 85]. In order to return agency to the user, design interventions, design friction, or nudges have been discussed as potential counter-measures to mitigate loss-of-control feelings getting users to reflect on their choices [40, 53, 84]. Focusing on the social medium Facebook, Lyngs et al. [55] compared two design interventions to empower users to reflect on their usage behaviour. One would periodically remind the users about their initial goals; the other removed the newsfeed feature from Facebook to keep users' on track with their goals. In a controlled setup, the study revealed certain shortcomings but highlighted how interventions can help users reflect on their behaviour. Exploring how nudges can be used similarly, Masaki et al. [57] developed nudging interventions deployed to shield users' privacy. These findings gain support by work conducted by Wang et al. [83], demonstrating the positive effects of design interventions, friction, and nudges that can grant users better agency when engaging with social media. In a similar vein, Lukoff et al. [54] conducted

a co-design study based on YouTube’s mobile interface to develop alternative mechanisms that redirect a sense of agency to the platform’s users. The limited agency of social networking service (SNS) users could be an indicator of decreased well-being. While previous studies promote design interventions as counter-measures to problematic interfaces in this regard, we return to traditional UCD principles aiming to understand users’ expectations regarding UI features within and beyond the SNS context. Exploring the possibilities for adaptive interfaces users can customise to meet their needs, Kollnig et al. [51] developed an “app repair framework”, after a majority of 85% of their participants demonstrated appreciation for the option to alter elements of their apps. With a focus on audio-related privacy when recording, Dunbar et al. [23] propose design principles to accommodate people’s concerns and preferences in this sensitive context. These examples highlight the necessity of understanding when and how in the design process to factor in considerations regarding user agency.

Inspired by this body of work, our objective is to extend our focus beyond specific usage contexts and conceptualise the nuances of design choices practitioners encounter. Furthermore, we intend to incorporate ethical considerations from dark pattern scholarship into the discourse at a more general level. To that end, we connect the fields of cognitive biases and dark patterns to explore their relationships that often lead to negative and harmful effects for users.

2.2 Cognitive Biases

Utilisation of human cognition and perception is an integral part of human factors and HCI [86]. In this vein, Jain and Horowitz et al. [48] discuss opportunities that go beyond interactions requiring users to make conscious decisions. As the work rests its focus on HCI-related work to promote seamless, unconscious interactions that could enhance people’s experience with novel technologies, they remind of ethical constraints when designing to alter cognitive processes. While our work is inspired by underlying constructs of cognitive biases, governing our choices unconsciously, the overall history of this field exceeds the means of this paper. Here, we synthesise key contributions as a background of our work.

Tversky and Kahneman’s [50] first introduced the term cognitive biases in 1972. Ever since, related work has identified and described a plethora of effects that influence human decision-making. The granularity and diversity of cognitive biases are well reflected in Baron’s book “Thinking and Deciding” [6], highlighting the enormous effort that went into this field of research, reporting that the variety of concepts has been thoroughly catalogued and understood. In an attempt to summarise these past efforts, Hilbert [47] identified eight core biases, provoking the idea that every other cognitive bias can be mapped to one of them. Hilbert placed these core biases alongside additional mathematical definitions to avoid any future ambiguity. His framework can help to understand the logical constraints in cognitive processes and promotes the concept of “The Noisy Memory Channel,” demonstrating how noises — confusion and mistakes — affect decision-making.

With the aim to help people make informed decisions, Hertel et al. [45] present a comprehensive overview of cognitive bias modification (CBM) strategies, which encompass procedures designed to prohibit automatic cognitive processes. This includes the alteration of peoples’ attention, interpretation of situations, and their memory to affect future decisions. Based on the growing body of work investigating CBM, Jones et al. [49] conducted a systematic literature review. The authors demonstrate the effectiveness of CBM across selected studies, which consistently show that CBM can modify targeted biases in adults but not children. This work underlines the potential of CBM as a tool for influencing automatic decision processes and reducing cognitive biases. Between the lines, these works postulate certain negative connotations and consequences of cognitive biases. Haselton et al. [41] spotlight certain situations where the opposite occurs. Their work on Error Management Theory (EMT) asserts that biased judgments in uncertain conditions can actually result in better decisions compared to unbiased alternatives as part of ongoing research studying the evolution of cognitive biases [42].

While these different perspectives mirror the power with which cognitive biases steer decisions, work illustrates the ease in which they can be exploited, for instance, to personalise recommender systems [77] for increasing purchasing behaviour or govern peoples’ search behaviour [5]. To explore possibilities to account for cognitive biases to avoid harmful designs, work by Kahneman and Tversky [78] and Baron [6] informed the design of our focus groups. We relied on the general definitions and selected biases from these works as part of each focus group. We thus base our findings on the fundamentals of cognitive bias and heuristic decision-making theory to learn about their relation to dark patterns.

2.3 Dark Patterns

Critique on nudges [40] spotlights ethical design caveats to obfuscate people’s decision-making. In the scope of HCI, these concerns have sparked various streams of research to investigate the negative or harmful effects of technological interactions — in its midst a discourse surrounding dark patterns. Introduced by Brignull in 2010 [11], the concept describes unethical practices that trick users into undesired actions with negative or harmful consequences. Alongside this discourse, research in HCI has fostered a growing typology of unethical graphical user interfaces fitting the definition of dark patterns across digital media. These include, but are not limited to, games [87] and social media [64, 71], as well as context-specific instances where language barriers are exploited [46]. This effort has recently led to the development of a first ontology of related design patterns [30]. *Interface Interference* [32], for instance, promotes certain interface elements, often visually, to gain users’ attention with instances identified in e-commerce [58] and social media [63]. Collectively, these works demonstrate the multitude of deceptive and manipulative strategies practitioners use to steer users’ decisions against their best interest and toward service providers’ goals. To offer an overview, Mathur et al. [59] constructed overarching characteristics by reviewing over 82 types of strategies that include research and regulation-based types. Mildner et al. [64] studied the

effects of 80 empirically studied types in the context of social media, while Gray et al. [30] considered a corpus of over 245 for their ontology.

While studies have frequently proposed a relationship between dark patterns and cognitive biases [58, 82], to our knowledge, we are among the first to explore this relationship in a dedicated study. Allowing interpretation in this vein, various studies convey difficulties among their participants to effectively recognise and avoid the effects of dark patterns [10, 22, 56, 63]. Aiming to create privacy-protecting interventions, work has deployed design strategies as so-called “bright patterns” [36] that invert the mechanisms of dark patterns. The study demonstrates an arguably positive impact on users’ decisions. However, the study also shows that users’ choice architecture is similarly altered as it would be by dark patterns and, thus, restricting users’ autonomy to make informed decisions. This aligns with research conducted by Ahuja and Kumar [2], who demonstrate how dark patterns restrict user autonomy on various levels, bridging the important topics. As the overall goal of this paper is to understand the underlying mechanisms of autonomy-respecting design, we build on their efforts and focus on cognitive and behavioural aspects, particularly nudging and cognitive biases.

2.4 Terminology

Recent voices within the dark pattern community have argued that contemporary terminology — “dark pattern” — should not be used as it poses potential racial misconception [1]. In an attempt to offer an alternative, Brignull [11], who originally coined the concept, suggested using “deceptive design” to describe the unethical design practices. However, voices objected to this change [68], arguing the term “dark” does not insinuate a “bad” interaction but suggests hidden consequences. Furthermore, it dismisses a connection to pattern language [3], which conceptualises reusable design strategies and offers solutions to similar problems. This has led researchers to come up with different terms [66] to describe the same concept. While this discourse still unfolds, we opt to use the term “dark patterns”, staying coherent with related work placing its origin in *hidden consequences* of interfaces instead of evilness or mal-intent, but acknowledge the importance of the current discussion problematising the term.

3 METHOD

To answer our research question and conceptualise the dynamic relationship between cognitive biases and dark patterns, we conducted a focus group study with experts in dark patterns and cognitive science/psychology research. Several participants in the study also bring extensive experience in interaction design, enriching our understanding of the practical implications in this field. We opted for expert participants to accommodate the exploratory nature of our research question. Additionally, the engagement of these two relevant perspectives in focused discussions promises novel and interesting insights into the similarities, differences, and facilitators of the two concepts. For each focus group, we invited two expert participants with extensive knowledge of dark pattern scholarship and another two with academic backgrounds in psychology or cognitive science. After agreeing to participate, we sent relevant study

information to each participant. At the same time, we gained their consent to record and analyse each session in line with the host university’s guidelines. Except for one focus group featuring three participants, each was attended by four experts, resulting in a total of 15 participants. To accommodate their international backgrounds and the different time zones they were living in, the focus groups were held online via the video conference tool Zoom.

3.1 Participants

Participants were recruited from the authors’ professional and academic networks or word of mouth. We carefully selected individuals with backgrounds and expertise in dark patterns and cognitive science or psychology with the precondition of having published in esteemed venues of their respective fields. Participation was entirely voluntary and without compensation. Before participating in the focus groups, participants were sufficiently informed about the study’s purpose and design, about their rights following GDPR guidelines and then asked to give their informed consent. In total, we recruited 15 participants, six self-identified as female, eight as male, and one as non-binary. The average years of experience participants had in their fields at the time of conducting this study was 6.73 years ($sd = 3.43$). On average, participants were 32.87 years old ($SD = 6.08$). At the time of running this experiment, their professions included (Assistant) Professors (5), Postdocs/Senior Researchers (5), and PhD candidates (5). The demographics of our participants are summarised in Table 1.

3.2 Focus Groups

In the course of 90 minutes, each focus group followed the same procedure: After a brief introduction of all participants, commonly used definitions of the terms *cognitive bias* and *dark pattern* were provided in the form of an online presentation via Zoom. To illustrate previously discussed connections between the concepts, we showed an example illustrating common streaming subscription tiers, including the ‘framing effect’ and ‘decoy effect’ to demonstrate their enabling of *aesthetic manipulation*, *visual interference* as well as *price comparison prevention*, *hidden information*, and *sneaking* dark patterns. Figure 1 shows the example given during this introduction.

Initial Discussion. Afterwards, the group was asked three questions to be discussed within ten minutes each: (1) What are the key similarities between cognitive biases and dark patterns? (2) What are the key differences between cognitive biases and dark patterns? (3) What does a cognitive bias facilitate to become a dark pattern, especially with regard to design? Each question affords participants to consider the topic from a different angle. The participants were prompted to use the communication channel of their choice to voice their opinions. They could make use of the chat function of the video conferencing platform or voice their perspective directly via the audio channel.

Card Sorting Task. After the initial discussion, participants were coupled with a disciplinary counterpart to conduct a card sorting task. Using the online collaboration tool Miro, both pairings² were

²For the group featuring three participants instead of four, we joined all experts within a single group to complete the card sorting task together.

PARTICIPANT TABLE						
ID	Age	Gender	Country of Residence	Education	Occupation	Years of Experience
P1	47	Male	Ireland	PhD	Senior UX Researcher	3
P2	32	Male	Germany	PhD	Postdoc	8
P3	40	Female	Netherlands	PhD	Assistant Professor	7
P4	25	Male	United States	MSc	PhD Candidate	3
P5	28	Female	United States	MSc	PhD Candidate	5
P6	28	Male	Finland	PhD	Postdoc	5
P7	29	Female	United States	MSc	PhD Candidate	6
P8	29	Female	Germany	Diploma	PhD Candidate	4
P9	39	Non-binary	United States	PhD	Professor	15
P10	31	Male	Ireland	PhD	Postdoc	6
P11	32	Male	Germany	PhD	Postdoc	7
P12	40	Male	Ireland	PhD	Assistant Professor	10
P13	29	Female	Luxembourg	MSc	PhD Candidate	3
P14	29	Female	United States	PhD	Assistant Professor	12
P15	35	Male	Finland	PhD	Assistant Professor	6
<i>Mean</i> = 32.87			<i>Mean</i> = 6.73			
<i>SD</i> = 6.08			<i>SD</i> = 3.43			

Table 1. This table presents an overview of expert participants of the four focus groups.

then tasked to group selected cognitive biases and dark patterns within 20 minutes. To afford the timely constraints of each focus group, we chose to select only prominent cognitive biases and dark patterns based on the citation counts of respective publications. Because of this limitation, however, we did not analyse these results further to avoid a potentially strong selection bias. To give participants more context, cards included definitions taken from original publications. Despite its limitations, we included this exercise as we wanted to engage participants in a transdisciplinary activity to experiment and test previously discussed ideas. Further, our aim was to inspire alternative perspectives and enrich the following discussion. The supplementary material contains a snapshot of the initial Miro board, including the selected cognitive biases (sourced from Tversky’s work [78]) and dark pattern types (sourced from the dark pattern ontology by Gray et al. [30]).

Reflective Discussion. Following up on the card sorting task, participants discussed their experiences, difficulties, and ideas behind creating groups. In addition, the experience gained in the previous activity led to new insights with regard to the questions raised in the initial discussion. This allowed a more critical reflection on earlier statements as well as identifying links between cognitive biases and dark patterns that had gone unnoticed before.

4 ANALYSIS

Once we completed the four focus groups, we manually transcribed the discussions and prepared the data for further analysis. To this end, we anonymised any content that could be traced back to any individual participant. For the analysis of the transcribed data, we conducted a thematic analysis and used the card sorting results as an assistive source for a better understanding of participants’ perspectives. These steps were concluded using the software Atlas.ti [29].

4.1 Positionality

The authors of this research have mixed backgrounds. One author gained their education in West Africa while the others acquired their education in Central Europe with WEIRD (Western, Educated, Industrialised, Rich, and Democratic) [44] backgrounds. Their research expertise includes design, computer science, and psychology, while their academic work focuses on topics concerned with social justice and user well-being in digital technology contexts. The focus groups were transcribed by two authors and, afterwards, coded by two authors. All participants were recruited through professional networks and by word-of-mouth. Each focus group was conducted by at least two authors, one of which would be responsible for moderating, the other(s) held assistive role(s). As this research aims to describe the relationship between cognitive biases and dark patterns, this lens has guided the structure and analysis of the focus groups. Finally, we acknowledge any possible biases that are the result of our academic, cultural, and personal backgrounds.

4.2 Coding of the Focus Group Transcripts

Prior to the analysis, we manually transcribed and anonymised the focus group discussions. In the first step, two authors coded a representative sample of 50% of the material using open coding in line with Blandford et al. [8]. We then conducted an iterative discussion to establish an initial coding tree. The remaining transcripts were split between the two authors and coded individually. Finally, we conducted a concluding discussion session to finalise the coding tree. This was followed by a thematic analysis to identify emerging dimensions from the material as described by Blandford et al. [8]. The codebook used to analyse the transcripts of the focus groups comprises 52 codes and is included in this paper’s supplementary material.

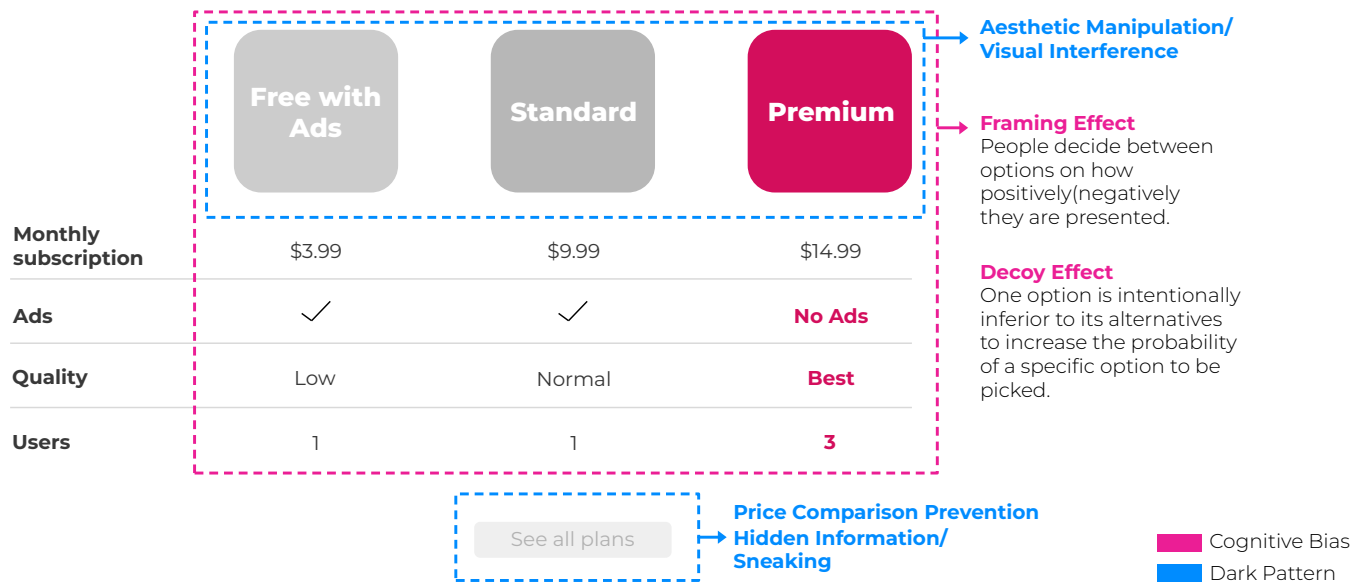


Fig. 1. This Figure was part of the introduction in each focus group and illustrates a common monthly subscription model of service providers with different tiers. The figure highlights in magenta the presence of cognitive biases (framing effect and decoy effect) and in blue dark patterns (aesthetic manipulation/visual interference, price comparison prevention, and hidden information/sneaking).

The main purpose of the card sorting task was to engage participants in an interactive task where dark pattern scholars and psychologists/cognitive scientists worked together to engage with and analyse their ideas from previous discussions. Hence, the set of biases and dark patterns used for the card sorting task was limited and only included subsets of the overall cognitive bias and dark pattern typologies. Thus, results were used to build an understanding of the data collected in the different discussion phases, therefore aiding our analysis of the focus group transcripts and the following identification of relationships.

Using axial coding, two authors began to independently organise all codes systematically into hierarchical structures and groups [8, 20]. This process was extended by connecting individual codes to describe their relationships, noting whether any confirming or contradicting notions exist between them. At this stage, the authors frequently revisited transcripts and card sorting results to ensure they stayed truthful to the data. Afterwards, the authors exchanged their findings in a following discussion. Here, the emergent hierarchical frameworks of the codes were the focus of identifying overarching themes. We thereby aimed to identify phases that followed design practice to real-world consequences in terms of dark patterns. Consequently, based on the analysis outlined above, we constructed a model that encapsulates these phases, effectively bridging the gap between design practices and their tangible impacts in the context of dark patterns.

5 FINDINGS

Based on the thematic analysis, we gained certain insights into the relationship between cognitive biases and dark patterns. However,

this relationship appears intricate and, expanding prior suggestions [58, 82], multifaceted. Here, we echo our participants' discussions to provide common and specific characteristics of the two fields. Moreover, we focus on the implications of design, the inscription of functionalities, and real-world applications as discussed by our participants, and promote the "Relationship Model of Cognitive Biases and Dark Patterns". Our presentation of the results as a model is consistent with established practices of presenting results in HCI and Ubicomp (e.g. [23, 25, 48, 62]).

5.1 Similarities, Differences, and Facilitators

Each focus group included granulated discussions about similarities, differences, and facilitation between cognitive biases and dark patterns. Although the participants either had backgrounds in dark pattern scholarship or cognitive science/psychology research, combined with experience in interaction design, providing definitions and a visual example established a common ground for fruitful discourse.

5.1.1 Similarities. When asked about the similarities between the two topics, participants noticed shared attributes regarding the impact on decision-making and autonomy. Arguing from a designer's perspective, P5 discussed this sentiment further when persuading user actions:

"I think the key similarity is [...] how the steering happens. [...] You are trying to steer your users from the normative ways of user agency, which is what HCI design mostly preaches." – P14

By leveraging users' perception and cognition, responsibility plays a pivotal role in the profession of designers. In this regard, P3 noticed another similarity in the shared dangers and harms that occur when not cared for. The same participant went on to discuss the particularities of users' unawareness:

"[B]oth can be unconscious. So, in both cases, the user may not be aware of the influence on their decision-making."
– P3

Overall, participants noticed certain synergies between cognitive biases and dark patterns. When triggered, both carry risks of making unfavourable choices as either is difficult to avoid. Certain responsibility was further attributed to designers utilising cognitive biases when used to steer users against their will.

5.1.2 Differences. The second question sought to identify differences between cognitive biases and dark patterns. Across all focus groups, a dominant argument addressed the different natures of either concept. While *"biases are already there [...] and inform our decisions"*, as P9 pointed out, or *"your intuitive brain [...] working in free-flow without really engaging too much in reasoned thought"* (P1), dark patterns, on the other side, are actively created and deployed by practitioners. P1 later continued their argument in consideration of the power dynamic between the designer and user:

"[A] designer has complete control over how that pattern takes form, takes shape, and how it's implemented within a system. Whereas a user doesn't have control over their cognitive biases." – P1

As cognitive biases are intrinsic to our behaviour, dark patterns, like all design patterns, are created and impact our behaviour intrinsically. Importantly, participants noticed that not all dark patterns actually require cognitive biases. This is in line with a statement made by P10:

"[O]bfuscation and sneak-into-basket and hidden costs, all of those just seem like outright lying. So, it's not necessarily using a cognitive bias to be the problem." – P10

5.1.3 Facilitators. During the last question, participants explored possible facilitators between the concepts. As similarities and differences foreshadow related attributes, P9 discusses how dark patterns emerge:

"[D]ark patterns are really the design material that allows those cognitive biases to be activated [...] this method is used to take advantage of our cognitive biases." – P9

A strong sentiment across participants for a facilitator was further noticed in (mal)intent behind deploying dark patterns. In this regard, P3, P12, and P15, thoroughly discussed designers' roles in persuading users' decisions toward commercial goals, limiting their autonomy. Sharing this position, P10 noted:

"Something is dark patterns when you use any means whatsoever [to] deprive me of my autonomy or to [...] engage me in practices that violate my privacy." – P9

Although not all dark patterns rely on cognitive biases, the latter seems to be an effective means to the end for malicious strategies. This reconnects to the previously discussed responsibilities of designers but also highlights additional needs for ensuring user safety.

As P3 anecdotally pointed out, however, regulations, such as the GDPR, do not consider intent as a necessity. This would ensure user protection whenever harm is done.

5.2 The Relationship Model of Cognitive Biases and Dark Patterns

To address the echoed implications for responsibility and impact for practitioners, we opted for developing a model that follows design from its creation to real-world implications. In a preliminary version, we modelled how cognitive biases and their exploitation can lead to deceptive design. Our model was constructed based on the findings of our study juxtaposed with relevant previous work. We verified our model by inviting focus group participants to provide feedback on individual and general levels of this preliminary model, helping us iterate and improve it where necessary. To this end, we sent out an online survey to all participants, informing them about the task and gaining their consent before breaking down the model for individual critique. Additionally, we offered participants to reach out to us and share further comments. In total, four participants responded to the survey, helping us to advance the model.

Based on the focus groups, collected feedback, and prior work, we introduce the *Relationship Model of Cognitive Biases and Dark Patterns* as an answer to our research question (visualised in Figure 2): How can we conceptualise the dynamic relationship between cognitive biases and dark patterns? This model comprises three stages that encompass five phases, showcasing a process from design to the real world. More precisely, inspired by ethical considerations of dark patterns [34] and their relationship toward cognitive biases [58, 82], the model follows the implementation of dark patterns (or other deceptive designs) from addressing cognitive biases in design to potential real-world implications, thereby showcasing potential harmful consequences which require the assessment of responsibilities.

Drawing inspiration from Verbeek's theory of technology mediation [80, 81], our model comprises three stages: (1) Inscription / Delegation from a designer's perspective, (2) mediation of the technology or particular interface, and (3) users' interpretation thereof. Adopting Verbeek's theory, the model demonstrates how dark patterns emerge from exploiting cognitive biases. To this end, our model breaks the three stages down into five phases. The first stage describes the designer's perspective to inscribe or delegate functionalities. Happening in phases one and two, **design addresses particular cognitive biases** determining the **balance between autonomy versus coercion**. The second stage – mediation – contains the third phase: the **exploitation of cognitive biases**, which can ultimately lead to deceptive practices and harm. The third stage focuses on the users' point of view, interpreting the design throughout the fourth and fifth phases. First, users **experience the designs' implications**, leading them to **questioning responsibility**. Additionally, we identified crossroads for safeguarding strategies of end-users as well as opportunities for organisations to limit harmful design through exploiting cognitive biases. In the following subsections, we outline each phase in detail. We support the descriptions for each phase through quotes from our participants and connect individual phases to related work where applicable. For improved

readability, we slightly altered some statements, ensuring words and sentiment were maintained.

5.3 From Design to Real World

The aim of the *Relationship Model of Cognitive Biases and Dark Patterns* is to follow the development of (un)ethical design from its planning stages until its deployment into the real world. Importantly, as with many things, breaking decisions and their consequences into stages or phases can result in over-simplifying processes. Our model is no exception. To emphasise the responsibility tied to designers' decisions and connect them with the consequences that arise once a design is implemented in the real world, we represented this progression with a continuous arrow beneath the three stages. Research [32, 59] and regulation and legislation [9, 52, 69] suggest that malintent plays a role in the design of online interfaces to deceive or manipulate users. However, many design decisions are not aimed at exploiting users' cognitive biases, and neither are practitioners always able to predict resulting harms. Throughout development, many constraints (e.g. money and time) limit the possibilities in which a design [31, 80] can be tested. Over-simplifying these issues can deflect from the critical aspects when design turns deceptive. Echoing the voices of our participants, the following phases attempt to convey these critical aspects with the goal of respecting the underlying continuum.

5.4 Phase 1: Design Addresses Cognitive Biases

The first phase of this model addresses the utilisation of cognitive biases to control users' attention or afford specific interactions. This is in line with Cross' description of design cognition [21], where it is part of the design process to identify the right problem to be solved: Depending on the situation and the underlying goal, a design may draw the user toward or away from it. Interactions are either better supported or actively obfuscated to guide users' decisions. While there are plenty of reasons for either strategy, dark patterns are used to provoke certain choices. It is important to emphasise here that our data show that specific cognitive biases are addressed both consciously (conscious decision of the designer) and unconsciously (not deliberately addressed by the designer). The *Bad Default* dark pattern [13], for instance, is often used in privacy settings where the system provider is initially allowed to collect personal information until a user decides otherwise. This pattern finds support in the 'default effect' as people often follow existing choices [6]. Whether the deployment of these strategies and usage of supportive cognitive biases is a deliberate choice is questioned by P14:

"Are [practitioners] doing this knowingly? I mean, the obvious is that [if] you try to get people to buy a certain thing, [...] how dark is a dark pattern if it's just accidental?" – P14

Designers can address the *intrinsic* nature of cognitive biases – shared between humans – when creating or using dark patterns to alter peoples' choice architecture. The particular relationship between intrinsic cognitive biases and extrinsic deployment of dark patterns was repeatedly discussed among participants across the focus groups. P15 and P12 populated this idea through the following statements:

"[C]ognitive biases are also so embedded in our automatic decision-making process, while dark patterns are our techniques deployed by external entities. [...] One is individual, embedded decision-making, and the other is external to us." – P15

"[C]ognitive biases are the receiver while the [dark pattern] are something from the perpetrator." – P12

The extrinsic effectiveness of dark patterns thereby hugely benefits from intrinsic cognitive biases, described by some participants as a symbiotic relationship. The knowledge garnered by cognitive scientists and psychologists plays a large role in informing dark patterns and spotlighting opportune moments to leverage cognitive biases. However, designers are not always in possession of this knowledge. P15 elaborated on the potential of this knowledge by saying:

"Knowing how we process information and how we heuristically process and think is the way dark patterns become so effective." – P15

5.5 Phase 2: Balance Between Autonomy vs. Coercion

In the second phase, designers navigate the challenging equilibrium between empowering users by giving them autonomy over their choices and deceptive practices that coerce users' decisions. The (conscious and unconscious) decisions made in the first phase have a direct influence on this balance, where ethical considerations and consequences are introduced. As with the first phase, the balance between providing users with autonomy and coercing their actions can be an active design choice or an unaware consequence [17]. The effect on the user, however, is often the same. Therefore, this phase requires careful attention to support reflective and informed decision-making. P10 argues that it can be dangerous for users when people attribute expertise to themselves in areas they do not actually have and that this can potentially cause a lot of harm, especially in sensitive contexts such as mental health applications:

"[S]elf assigned expertise [...] gives [practitioners] that right framing that they could then use dark patterns that exploit cognitive biases to design products without feeling like they are being very unethical." – P10

Instead of designing responsibly with users' best interests in mind, common or best practices and traditional mindsets within professions may foster excuses for harmful design implications at the user's cost [17, 31]. In this regard, the 'status-quo' bias [6] may offer some insight into why practitioners and designers resort to deploying the same problematic designs. As with persuasive design and nudges, however, their intention might be noble, but the outcome is detrimental to users' well-being. The plethora of deployed strategies that harm users [30] is witness to practitioners' negligence in this regard, reflected by the increased regulatory efforts in place to protect users [52, 69]. However, it has to be noted that this negligence might stem from a culture of oversight or neglect at the organisational level [17]. This second phase concludes the first stage of our model and the perspective of designers who inscribe their ideas into the design.

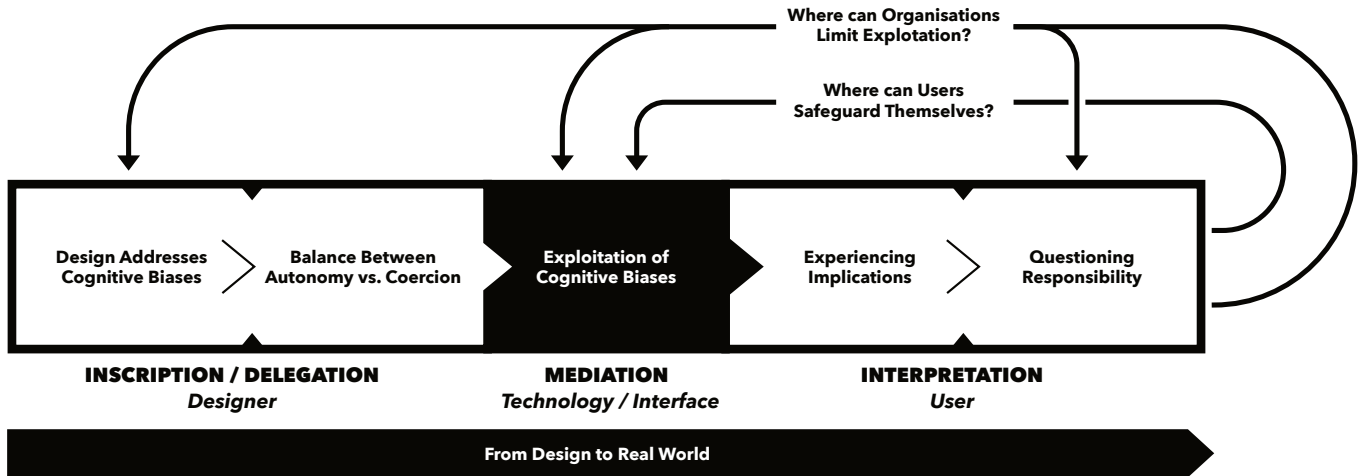


Fig. 2. This Figure presents the *Relationship Model of Cognitive Biases and Dark Patterns*. Following a continuum from (potentially unethical) design to real-world applications, the model comprises three stages spanning five phases. Adopting Verbeek’s theory of technology mediation [80, 81], the model follows designers’ inscription of functionalities into technology to users’ interpretation, leading to the questioning of responsibilities. Depending on the impact and implications of the (unethical) design, end-users may need safeguarding measures, while policy and regulation may be required for their protection.

5.6 Phase 3: Exploitation of Cognitive Biases

Relying on practitioners’ decisions in the first stage, the second stage contains the mediation of the design and the third phase where exploitation of cognitive biases manifests — eventually resulting in deceptive practices and harm. It conveys the implications of design that harnesses cognitive biases and controls user behaviour. In other words, this phase is about users’ interactions with the design and, thereby, the first phase, where design enters the real world with all its implications. It describes the users’ reaction to the design but not yet their evaluation of any interaction. Dark patterns that govern user interactions influence their choice architecture without providing transparent information about their consequences. In this vein, P9 drew the following connection:

“[A] cognitive bias is a necessary component as kind of [...] an ingredient to a dark pattern”. – P9

A core element to enable design — whether for good or bad — is to afford specific interactions [67]. To that end, affordances can have many forms and shapes to convey their aims but have to be delivered easily accessible for users to engage. A common choice to create affordances is through nudges [76]. While work demonstrates their effectiveness [14], they limit transparency as P9 detailed:

“[N]udges for good fail to respect the autonomy of the person who’s using the system if they’re not very open and transparent about how the nudges are working and there becomes a power imbalance then between whoever is providing those nudges and the person on the other end of them”. – P9

This power imbalance is a critical implication of dark patterns as the user is rarely in control or aware of any consequences [63]. Thereby, design can easily facilitate its goals without providing sufficient information that would allow users to reflect on their

choices. P10 explained how preferences can be exploited in the form of cognitive biases:

“[U]sing our tendency to prefer colour or be attention-grabbing so that we look at [the interface] more or our tendency to kind of not want to think more than we have to. [...] [D]ark patterns [are] more of a method and cognitive biases [are] more of an inherent thing that is exploited.” – P10

Our participants also discussed the natural benefits of certain cognitive biases and heuristics to provide effective shortcuts in various situations. In tandem with current design practices and service providers’ monetary incentives [16], their mere presence opens the door for serious exploitation, limiting users’ ability to formulate reflected decisions and avoid manipulation into unfavourable interactions.

5.7 Phase 4: Experiencing Implications

The fourth phase of the model addresses the real-world implications behind dark patterns and how users experience them. Here, the third and last stage begins, where users interpret the design. At this time, the impacts of the design become noticeable and perceptible, whether originally accounted for or unintentional. Informed by the first two stages of the model, during the fourth phase, practitioners have already deployed the strategies to deliver their design’s goal(s). This includes any decisions that affect user agency and autonomy through cognitive biases. In other words, one could say that from a user’s perspective, the harm is done. Elaborating on users’ expectations, P10 raises important questions from a user’s point of view:

“[W]hat mental models do I have about these products? And in what ways [do] my mental models of this product align with what’s happening before me?” – P10

Before any interaction, users may have certain expectations of what will happen. The previous phases and underlying choices ultimately lead to the moment when a user interacts with a design for the first time. While certain dark patterns prohibit reflective thinking, others obscure access to specific functionality. Unaware of the practitioners' goal, users rely on a design's attributes to understand required actions. Thereby, they can draw from their mental model and past experience before engaging with it but cannot know for certain what the consequences will be. At this stage, practitioners and users will only be able to assess the truth behind the implications once harm has been done.

5.8 Phase 5: Questioning Responsibility

In response to the aftermath of the prior phase, the fifth phase mirrors past decisions and reflects the question of responsibility behind harmful implications, concluding the third stage of interpretation. However, this question is not simply answered. From a user's perspective, technological illiteracy or unawareness of the presence of dark patterns can lead to seeking responsibility in themselves. Practitioners, on the other end, may displace responsibility behind best practices and common, albeit critically viewed, design paradigms or deflect responsibility altogether, convinced of the noble cause of their design. In any case, the responsibility is difficult to ascribe to a single party. Nonetheless, there are multiple avenues to mitigate negative implications. In this regard, P1 reflected on unintended but harmful interactions, taking the practitioners' perspective:

"[T]here is some kind of control over it from a designer's point of view that once [the harm] has been identified and they've observed the problem, that they can go back and fix it then afterwards". – P1

However, the strategy outlined by P1 would require a thorough understanding of exploited cognitive biases, alternative strategies, and access to necessary resources to change the design post-development. Especially in this phase, it is important to remember this model also as a continuum where decisions are intertwined and causally dependent. Hence, untangling the root causes behind implications and identifying responsibility is difficult. Therefore, practitioners need to acquire a critical view of their own work and reflect on utilised strategies to understand how cognitive biases enable dark patterns that affect end-users' decision-making. This does not mean that practitioners need to become experts in human cognition and perception (although it would not harm). Reflective steps throughout development stages can already help ensure that user autonomy is sufficiently respected to empower informed decision-making.

5.9 Safeguarding and Counter Measures

Collectively, the five phases of the *Relationship Model of Cognitive Biases and Dark Patterns* highlight how dark patterns leverage cognitive biases and how the latter can enable the first. However, counter-measures to existing dark patterns and their harms have not yet been accounted for. We, therefore, expanded the model with two additional arrows that link to opportune moments where users could safeguard themselves and where organisations, such as governmental bodies, could restrict the exploitation of cognitive biases

through guidelines and regulations. Suggesting that knowledge can help users mitigate harmful effects, P9 stated the following:

"Sometimes, if it's raised to your attention, you can fight back yourself. But in other cases, the pull is so strong that you actually need regulatory bodies or other kinds of policies in place to fight back on your behalf." – P9

Safeguarding measures for and by end-users are only effective if the problems are understood and can be avoided. In this regard, one arrow leaving the fifth phase carries the question: *Where can users safeguard themselves?* – points to the third phase, where cognitive biases are exploited. Users can learn to mitigate cognitive biases, for instance, through understanding how they manifest [18, 45]. But even then, it may not be possible to always rely on the safeguarding measures as demonstrated by Bongard-Blanchy et al. [10] or Mildner et al. [63]. Importantly, our participants have emphasised that this is not (ever) the user's fault. Hence, our participants envisioned that knowledgeable people would be able to recognise the potential exploitation of cognitive biases in the third phase and could proceed strategically with an appropriate reaction. P9 later continued their argument by discussing the current approach behind dark patterns and proposing legal requirements to assure user protection:

"I think a lot of the mainstream dark patterns that seem to be very effective at triggering our cognitive biases and altering our ability to act or understanding what the situation is, do certainly seem like they're illegitimate forms of altering sense-making versus ones that we would describe as normatively acceptable." – P9

In this regard, P3 points toward already existing regulation, such as the Digital Service Act (DSA) [69] of the European Union (EU), restricting the implementation of dark patterns. Exploitation of cognitive biases, however, is not regulated:

"There [is] the word dark patterns in the law now, but cognitive biases are not a term for law." – P3

In our model, organisational limits to exploitation, such as guidelines or regulations, are alternatives to end-user safeguarding. In this regard, the model incorporates a second arrow leaving the fifth phase asking: *Where can organisations limit exploitation?* The arrow links to the first, third, and fifth phases to highlight opportunities to protect users. In the same direction as pointed out by P3, organisations could request user autonomy to be respected, as done by the EU's DSA [69]. They could also restrict the exploitation of cognitive biases through regulating deceptive design, also done by the DSA as well as other regulations such as the US state of California's CCPA [52] or India's recent issue to prevent dark patterns [65]. Lastly, the arrow loops back into the fifth phase to illustrate the importance of identifying offenders and holding them accountable if users are harmed.

Although we view these measures as more effective compared to users' safeguarding themselves, we recognise that they are also more drastic and finite in the case organisations increase regulations. On the one hand, they could address the exploitation of certain biases directly, for example, by requiring specific user consent or autonomy to prevent harmful, impulsive interactions. This line of defence leaves space for practitioners to choose alternative design

strategies. On the other hand, regulations can restrict the overall available design space, prohibiting the deployment of dark patterns. As with the DSA [69], specific strategies can be directly addressed to ensure user safety.

6 DISCUSSION

Inspired by previous research proposing a connection between cognitive biases and dark patterns [58, 82], this paper offers answers to our research question by problematising the dynamic and intricate constraints within this relationship. To that end, we conducted four focus groups and, based on our results, proposed the *Relationship Model of Cognitive Biases and Dark Patterns*. This model depicts five phases that describe how decisions to deploy cognitive biases can enable exploitation and dark patterns. In this section, we discuss applications of this model, how technology can be devised to preserve user autonomy in such contexts and point toward future avenues HCI could consider to foster the implementation of ethical design and user protection.

6.1 Using the Relationship Model of Cognitive Biases and Dark Patterns

Previous work has described a growing typology of dark patterns [30, 59, 64] users constantly encounter across web and app interfaces. While various patterns have been captured and are relatively well understood, this strand of research lacks in-depth knowledge of the underlying mechanisms. Contributing relevant insights in this regard, the main aim of the *Relationship Model of Cognitive Biases and Dark Patterns* is to support both researchers and practitioners by providing them with reflective phases that can support them in considering the ethical caveats and impacts of their designs. It is not meant to be prescriptive but provides a guiding roadmap reminding about ethical implications throughout the design's lifespan and the interplay of cognitive biases and dark patterns in that regard. While the aim of our model cannot change any malicious objectives of practitioners, it can serve as a reminder about the implicated consequences of utilising cognitive biases and can guide toward potential counter-measures. The model can, therefore, be applied in situations where dark patterns are observed. **Especially in the early development stages of designs, the *Relationship Model of Cognitive Biases and Dark Patterns* can complement decisions made alongside existing, traditional design paradigms that may not always prioritize user agency and autonomy.**

The research community focusing on dark patterns [35] has recently questioned the intent behind deploying strategies that harm users. Throughout their discussions, participants across our focus groups separated practitioners' intent from the harmful impacts of deployed interactions. While some practitioners utilise obvious tactics to manipulate users' choices to their service's advantage, smaller service providers may naively follow existing common practices and deploy dark patterns without realising the consequences. However, in both cases, the end-user may suffer from possible implications, regardless of intent. Based on these insights, we argue that **necessary protective measures should be independent of intentions as they deem the safety of end-users most important.** Recent regulations [52, 69] already try to restrict certain dark patterns.

While this offers promising avenues to enhance user protection and safety, we maintain that regulatory measures should be employed judiciously when other protective measures have proven ineffective. The form and shape in which regulation should be enforced pose new, complex challenges. On the one hand, over-regulation may restrict innovation in the early stages. On the other hand, it is difficult to administer regulations that affect domains as large as web and app interactions.

6.2 Preserving User Autonomy

A key aspect of our model is to provide arguments for the importance of supporting and preserving user autonomy. Based on our findings, we argue that it is crucial to allow for and foster informed decisions before interactions happen. Traditional HCI paradigms offer practitioners a range of tools and common practices to support ethical design practice. Although critically reviewed, persuasive design explains how motivation can be directly addressed to steer and alter users' choice architecture and guide them toward a pre-determined goal. While behaviour change can be supported through design and technology, it is easy to mean good and do harm [12]. In research, it is customary to behave responsibly, ask for consent, and make users aware that the technologies utilised will potentially affect their choices. Outside research, users are often kept unaware of the consequences behind interactions, often noticed in infamous cookie-consent banners [34, 79]. Knowledge about the enabling effects of cognitive biases toward dark patterns could mitigate certain exploitation and protect users [49]. However, multiple studies [10, 22, 63] investigating the end-users ability to recognise and avoid dark patterns repeatedly demonstrate difficulties among their participants, even if information about dark patterns was provided. This indicates that user safeguarding is limited in its effectiveness, and other means are necessary to uphold their autonomy.

Overall, autonomy and agency-driven design have been a long-standing element in HCI [28]. However, a lack of fine-grained understanding of how to connect these principles with usability [53] suggests more work is needed to take full advantage of existing design paradigms. Our model invites researchers and practitioners to reflect on their responsibilities. We therefore hope that it can support continuing work and foster more responsible designs. Future work could study the effectiveness of the model alongside existing design methods. For example, it would be interesting to learn whether persuasive design or nudges can afford more autonomy if the model is used alongside such practices. Nevertheless, existing work in HCI echoes a balancing act between user autonomy and alternation of choice architecture in every design. **It remains within the designers' responsibility to understand this challenge and follow ethical design principles promoting user autonomy.** A future goal of this work could be the investigation of opportune design strategies that better connect usability, user experience, and autonomy and create better incentives and tangible examples for practitioners to follow such principles.

6.3 Ways Forward

Our proposed model mainly targets its depicted challenges from an HCI perspective. While the model offers novel insights in this

regard and is designed to assist practitioners in their work, it is only one step towards understanding the underlying mechanisms of dark patterns. P15, a participant with a psychology background, speculated into the direction of cognitive processing:

"I can come up with the whole information processing pipeline of human psychology here and just identify some [showing that] it's not covering everything." – P15

Although the discussions of our participants did not come to an overall conclusion or identify specific solutions, they provided rich food for thought about the different angles from which dark patterns could be studied. With more emphasis on human cognition, perhaps future work can work toward models that better support users' perspectives. An already existing angle in this regard is presented in cognitive bias modification (CBM). The work highlighting the effectiveness in which autonomous decisions can be altered to be better reflected [45, 49] spotlights an exciting direction to help users make better decisions when faced with dark patterns. Moreover, research in HCI could adopt CBM to devise technologies that prepare and shape users' expectations in line with possible interactions.

7 LIMITATIONS

Although we were careful in designing and conducting this study, we recognise that this work is prone to certain limitations. Firstly, the last focus group was attended by only three instead of four participants (two experts in dark pattern scholarship and one expert in psychology). While the focus group functioned well in terms of interpersonal dynamics, a fourth member might have changed the dynamic of the discussions. Moreover, this changed the dynamic in which prior card sorting tasks were conducted. Instead of two groups including one expert from the respective fields to discuss and execute the task, we decided to have one group including all three participants. Again, the outcome was comparable to prior focus groups, but we cannot know whether a second expert with expertise in cognitive science or psychology would have impacted the group's discourse and, thus, our findings.

Secondly, the study was designed and administered by three researchers with backgrounds in HCI research. Although one of them has a background in psychology, their current research is situated in the field of HCI. Although the study was informed by relevant work on the concepts of cognitive biases and human behaviour to ensure a levelled discussion in terms of the considered subjects, we acknowledge a certain bias stemming from our expertise. We would welcome any future attempts to reproduce our findings in the fields of human cognition and behaviour to gain different views that may result in additional findings.

Our personal HCI backgrounds also resulted in a third limitation regarding the *Relationship Model of Cognitive Biases and Dark Patterns*. Here, we recognise that our HCI background guided our moderation of the focus groups and informed our interpretation of the data as well as any further analysis. To mitigate these effects, we extensively reviewed material from the field of human cognition, particularly focusing on cognitive bias literature. However, researchers with other backgrounds may interpret the same data differently.

8 CONCLUSION

This research sheds light on the dynamic relationship between cognitive biases and dark patterns. Based on a focus group study with expert participants, we explore the ethical considerations throughout the development of design and describe the *Relationship Model of Cognitive Biases and Dark Patterns* providing an overview of the continuum between design and real-world implications of harmful design.

To that end, we emphasize the critical role of practitioners and researchers in considering the ethical implications of their design decisions, particularly when it comes to user autonomy. By illustrating the dynamic process through which cognitive biases are leveraged to create dark patterns, the model underscores the responsibility designers hold in shaping user experiences. Furthermore, the study recognises the importance of safeguarding strategies for end-users and regulatory measures to protect individuals from the potential harms of dark patterns. This recognition aligns with broader discussions in the HCI community regarding better implementation of ethical design practices and user-centred development of autonomy and agency-enabling technologies.

In summary, this research deepens our understanding of the intricate interplay between cognitive biases and dark patterns but also provides a practical model for practitioners to navigate this complex landscape responsibly. As technology continues to play an increasingly ubiquitous role in our lives, the ethical considerations highlighted in this study become even more pertinent as related work catalogues the variety of unethical, dark patterns. Thus, we hope that this research contributes to the ongoing dialogue on ethical design in HCI and sets the stage for future investigations into creating safer, more user-friendly digital interfaces.

REFERENCES

- [1] ACM. 2023. Words matter: Alternatives for charged terminology in the computing profession. <https://www.acm.org/diversity-inclusion/words-matter>
- [2] Sanju Ahuja and Jyoti Kumar. 2022. Conceptualizations of user autonomy within the normative evaluation of dark patterns. *Ethics and Information Technology* 24, 4 (Dec. 2022), 52. <https://doi.org/10.1007/s10676-022-09672-9>
- [3] C. Alexander, S. Ishikawa, and M. Silverstein. 1977. *A Pattern Language: Towns, Buildings, Construction*. OUP USA. <https://books.google.ch/books?id=hwAHmktpk5IC>
- [4] Dmitry Alexandrovsky, Maximilian Achim Friehs, Jendrik Grittner, Susanne Putze, Max V. Birk, Rainer Malaka, and Regan I. Mandryk. 2021. Serious Snacking: A Survival Analysis of how Snacking Mechanics Affect Attrition in a Mobile Serious Game. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Number 113. Association for Computing Machinery, New York, NY, USA, 1–18. <https://doi.org/10.1145/3411764.3445689>
- [5] Leif Azzopardi. 2021. Cognitive Biases in Search: A Review and Reflection of Cognitive Biases in Information Retrieval. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval (Canberra ACT, Australia) (CHIIR '21)*. Association for Computing Machinery, New York, NY, USA, 27–37. <https://doi.org/10.1145/3406522.3446023>
- [6] Jonathan Baron. 2007. *Thinking and deciding*. Cambridge University Press.
- [7] Dan Bennett, Oussama Metatla, Anne Roudaut, and Elisa D. Mekler. 2023. How does HCI Understand Human Agency and Autonomy?. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–18. <https://doi.org/10.1145/3544548.3580651>
- [8] 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.
- [9] 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.
- [10] 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. *Designing Interactive Systems Conference 2021* (June 2021), 763–776. <https://doi.org/10.1145/3461778.3462086> arXiv: 2104.12653.
 - [11] Harry Brignull, Marc Miquel, Jeremy Rosenberg, and James Offer. 2010. Dark Patterns-User Interfaces Designed to Trick People. <http://darkpatterns.org/> (visited on 2021-08-25).
 - [12] Hronn Brynjarsdottir, Maria Håkansson, James Pierce, Eric Baumer, Carl DiSalvo, and Phoebe Sengers. 2012. Sustainably unpersuaded: how persuasion narrows our vision of sustainability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Austin Texas USA, 947–956. <https://doi.org/10.1145/2207676.2208539>
 - [13] Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher. 2016. Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns. *Proceedings on Privacy Enhancing Technologies* 2016, 4 (2016), 237–254. <https://doi.org/10.1515/popets-2016-0038>
 - [14] Ana Caraban, Evangelos Karapanos, Daniel Gonçalves, and Pedro Campos. 2019. 23 Ways to Nudge: A Review of Technology-Mediated Nudging in Human-Computer Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland UK, 1–15. <https://doi.org/10.1145/3290605.3300733>
 - [15] Lorena Sánchez Chamorro, Kerstin Bongard-Blanchy, and Vincent Koenig. 2023. Ethical Tensions in UX Design Practice: Exploring the Fine Line Between Persuasion and Manipulation in Online Interfaces. (July 2023), 61–73. <https://doi.org/DOI:10.1145/3563657.3596013>
 - [16] Zhilong Chen, Jinghua Piao, Xiaochong Lan, Hancheng Cao, Chen Gao, Zhicong Lu, and Yong Li. 2022. Practitioners Versus Users: A Value-Sensitive Evaluation of Current Industrial Recommender System Design. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 533 (nov 2022), 32 pages. <https://doi.org/10.1145/3555646>
 - [17] Shruthi Sai Chivukula, Ike Obi, Thomas V Carlock, and Colin M. Gray. 2023. Wrangling Ethical Design Complexity: Dilemmas, Tensions, and Situations. In *Companion Publication of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) (DIS '23 Companion). Association for Computing Machinery, New York, NY, USA, 179–183. <https://doi.org/10.1145/3563703.3596632>
 - [18] Andy Cockburn and Carl Gutwin. 2019. Anchoring Effects and Troublesome Asymmetric Transfer in Subjective Ratings. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–12. <https://doi.org/10.1145/3290605.3300592>
 - [19] Anna L. Cox, Sandy J.J. Gould, Marta E. Cecchinato, Ioanna Iacovides, and Ian Renfree. 2016. Design Frictions for Mindful Interactions: The Case for Microboundaries. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) (CHI EA '16). Association for Computing Machinery, New York, NY, USA, 1389–1397. <https://doi.org/10.1145/2851581.2892410>
 - [20] John W. Creswell. 2009. *Research design: qualitative, quantitative, and mixed methods approaches* (3rd ed ed.). Sage Publications, Thousand Oaks, Calif. OCLC: ocn192045753.
 - [21] Nigel Cross. 2001. Design Cognition: Results From Protocol And Other Empirical Studies Of Design Activity. In *Design knowing and learning: cognition in design education*. Elsevier, Oxford, UK, 29–103. <https://oro.open.ac.uk/3285/>
 - [22] Linda Di Geronimo, Larissa Braz, Enrico Fregnan, Fabio Palomba, and Alberto Bacchelli. 2020. UI Dark Patterns and Where to Find Them. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–14. <https://doi.org/10.1145/3313831.3376600>
 - [23] Julia C. Dunbar, Emily Bascom, Ashley Boone, and Alexis Hiniker. 2021. Is Someone Listening? Audio-Related Privacy Perceptions and Design Recommendations from Guardians, Pragmatists, and Cynics. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 98 (sep 2021), 23 pages. <https://doi.org/10.1145/3478091>
 - [24] Salma Elsayed-Ali, Sara E Berger, Vagner Figueredo De Santana, and Juana Catalina Becerra Sandoval. 2023. Responsible & Inclusive Cards: An Online Card Tool to Promote Critical Reflection in Technology Industry Work Practices. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–14. <https://doi.org/10.1145/3544548.3580771>
 - [25] Daniel A. Epstein, An Ping, James Fogarty, and Sean A. Munson. 2015. A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) (UbiComp '15). Association for Computing Machinery, New York, NY, USA, 731–742. <https://doi.org/10.1145/2750858.2804250>
 - [26] BJ Fogg. 2009. A Behavior Model for Persuasive Design. In *Proceedings of the 4th International Conference on Persuasive Technology* (Claremont, California, USA) (Persuasive '09). Association for Computing Machinery, New York, NY, USA, Article 40, 7 pages. <https://doi.org/10.1145/1541948.1541999>
 - [27] Batya Friedman, Peter H. Kahn, Alan Borning, and Alina Hultgren. 2013. Value Sensitive Design and Information Systems. In *Early engagement and new technologies: Opening up the laboratory*. Neelke Doorn, Daan Schuurbijs, Ibo van de Poel, and Michael E. Gorman (Eds.). Vol. 16. Springer Netherlands, Dordrecht, 55–95. https://doi.org/10.1007/978-94-007-7844-3_4 Series Title: Philosophy of Engineering and Technology.
 - [28] Batya Friedman and Helen Nissenbaum. 1997. Software agents and user autonomy. In *Proceedings of the first international conference on Autonomous agents - AGENTS '97*. ACM Press, Marina del Rey, California, United States, 466–469. <https://doi.org/10.1145/267658.267772>
 - [29] ATLAS.ti Scientific Software Development GmbH. 2021. ATLAS.ti: The Qualitative Data Analysis & Research Software. <https://atlasti.com/>
 - [30] Colin M Gray, Nataliia Bielova, Cristiana Santos, and Thomas Mildner. 2024. An Ontology of Dark Patterns: Foundations, Definitions, and a Structure for Transdisciplinary Action. *arXiv preprint arXiv:2309.09640* (2024).
 - [31] 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*. ACM, Glasgow Scotland UK, 1–11. <https://doi.org/10.1145/3290605.3300408>
 - [32] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. The dark (patterns) side of UX design. *Conference on Human Factors in Computing Systems - Proceedings* 2018-April (2018), 1–14. <https://doi.org/10.1145/3173574.3174108>
 - [33] Colin M. Gray, Cristiana Santos, Nataliia Bielova, and Thomas Mildner. 2024. An Ontology of Dark Patterns: Foundations, Definitions, and a Structure for Transdisciplinary Action. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, 13 pages. <https://doi.org/10.1145/123456789>
 - [34] Colin M. Gray, Cristiana Santos, Nataliia Bielova, Michael Toth, and Damian Clifford. 2021. Dark patterns and the legal requirements of consent banners: An interaction criticism perspective. *arXiv* (2021). <https://doi.org/10.1145/3411764.3445779>
 - [35] Colin M. Gray, Cristiana Teixeira Santos, Nicole Tong, Thomas Mildner, Arianna Rossi, Johanna T. Gunawan, and Caroline Sindors. 2023. Dark Patterns and the Emerging Threats of Deceptive Design Practices. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–4. <https://doi.org/10.1145/3544549.3583173>
 - [36] Paul Graßl, Hanna Schraffenberger, Frederik Zuiderveen Borgesius, and Moniek Buijzen. 2021. Dark and Bright Patterns in Cookie Consent Requests. *Journal of Digital Social Research* 3, 1 (Feb. 2021), 1–38. <https://doi.org/10.33621/jdsrv.311.54>
 - [37] Barbara Grimpe, Mark Hartwood, and Marina Jirotko. 2014. Towards a closer dialogue between policy and practice: responsible design in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Toronto Ontario Canada, 2965–2974. <https://doi.org/10.1145/2556288.2557364>
 - [38] 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, CSCW2 (Oct. 2021), 1–29. <https://doi.org/10.1145/3479521>
 - [39] Juho Hamari, Jonna Koivisto, and Tuomas Pakkanen. 2014. Do Persuasive Technologies Persuade? - A Review of Empirical Studies. In *Persuasive Technology*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Alfred Kobsa, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Demetri Terzopoulos, Doug Tygar, Gerhard Weikum, Anna Spagnoli, Luca Chittaro, and Luciano Gamberini (Eds.). Vol. 8462. Springer International Publishing, Cham, 118–136. https://doi.org/10.1007/978-3-319-07127-5_11 Series Title: Lecture Notes in Computer Science.
 - [40] Pelle Guldberg Hansen and Andreas Maaløe Jespersen. 2013. Nudge and the Manipulation of Choice: A Framework for the Responsible Use of the Nudge Approach to Behaviour Change in Public Policy. *European Journal of Risk Regulation* 4, 1 (March 2013), 3–28. <https://doi.org/10.1017/S1867299X00002762>
 - [41] Martie G. Haselton and Daniel Nettle. 2006. The Paranoid Optimist: An Integrative Evolutionary Model of Cognitive Biases. *Personality and Social Psychology Review* 10, 1 (Feb. 2006), 47–66. https://doi.org/10.1207/s15327957pspr1001_3
 - [42] Martie G Haselton, Daniel Nettle, and Paul W Andrews. 2015. *The evolution of cognitive bias*. Wiley Online Library. 724–746 pages.
 - [43] Daniel M. Hausman and Brynn Welch. 2010. Debate: To Nudge or Not to Nudge*. *Journal of Political Philosophy* 18, 1 (March 2010), 123–136. <https://doi.org/10.1111/j.1467-9760.2009.00351.x>
 - [44] Joseph Henrich, Steven J Heine, and Ara Norenzayan. 2010. Most people are not WEIRD. *Nature* 466, 7302 (2010), 29–29.
 - [45] Paula T. Hertel and Andrew Mathews. 2011. Cognitive Bias Modification: Past Perspectives, Current Findings, and Future Applications. *Perspectives on Psychological Science* 6, 6 (Nov. 2011), 521–536. <https://doi.org/10.1177/1745691611421205>
 - [46] Shun Hidaka, Sota Kobuki, Mizuki Watanabe, and Katie Seaborn. 2023. Linguistic Dead-Ends and Alphabet Soup: Finding Dark Patterns in Japanese Apps. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*.

- ACM, Hamburg Germany, 1–13. <https://doi.org/10.1145/3544548.3580942>
- [47] Martin Hilbert. 2012. Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making. *Psychological Bulletin* 138, 2 (March 2012), 211–237. <https://doi.org/10.1037/a0025940>
- [48] Abhinandan Jain, Adam Haar Horowitz, Felix Schoeller, Sang-won Leigh, Pattie Maes, and Misha Sra. 2020. Designing Interactions Beyond Conscious Control: A New Model for Wearable Interfaces. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 3, Article 108 (sep 2020), 23 pages. <https://doi.org/10.1145/3411829>
- [49] Emma B. Jones and Louise Sharpe. 2017. Cognitive bias modification: A review of meta-analyses. *Journal of Affective Disorders* 223 (Dec. 2017), 175–183. <https://doi.org/10.1016/j.jad.2017.07.034>
- [50] Daniel Kahneman. 2003. A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist* 58, 9 (2003), 697–720. <https://doi.org/10.1037/0003-066X.58.9.697>
- [51] Konrad Kollnig, Siddhartha Datta, Thomas Serban Von Davier, Max Van Kleek, Reuben Binns, Ulrik Lyngs, and Nigel Shadbolt. 2023. 'We Are Adults and Deserve Control of Our Phones': Examining the Risks and Opportunities of a Right to Repair for Mobile Apps. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) (FAccT '23). Association for Computing Machinery, New York, NY, USA, 22–34. <https://doi.org/10.1145/3593013.3593973>
- [52] California State Legislature. 2018. CCPA-18 2018. California Consumer Privacy Act of 2018 [1798.100 - 1798.199] (CCPA). https://leginfo.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5
- [53] David Leimstädtner, Peter Sörries, and Claudia Müller-Birn. 2023. Investigating Responsible Nudge Design for Informed Decision-Making Enabling Transparent and Reflective Decision-Making. In *Mensch und Computer 2023*. ACM, Rapperswil Switzerland, 220–236. <https://doi.org/10.1145/3603555.3603567>
- [54] Kai Lukoff, J Vera Liao, James Choi, Kaiyue Fan, Sean A Munson, and Alexis Hiniker. 2021. How the Design of YouTube Influences User Sense of Agency. In *CHI'21*. ACM, Yokohama. <https://doi.org/10.1145/3411764.3445467>
- [55] Ulrik Lyngs, Kai Lukoff, Petr Slovak, William Seymour, Helena Webb, Marina Jirotko, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. 2020. 'I Just want to Hack Myself to Not Get Distracted': Evaluating Design Interventions for Self-Control on Facebook. In *CHI'20*. ACM, Honolulu, 1–15. <https://doi.org/10.1145/3313831.3376672>
- [56] Maximilian Maier. 2020. Dark Design Patterns - An End-user Perspective. *Human Technology* 16 (2020), 170–199. <https://doi.org/10.17011/ht/urn.202008245641>
- [57] Hiroaki Masaki, Kengo Shibata, Shui Hoshino, Takahiro Ishihama, Nagayuki Saito, and Koji Yatani. 2020. Exploring Nudge Designs to Help Adolescent SNS Users Avoid Privacy and Safety Threats. 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, 1–11. <https://doi.org/10.1145/3313831.3376666>
- [58] Arunesh Mathur, Gunes Acar, Michael J. Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–32. <https://doi.org/10.1145/3359183> arXiv: 1907.07032.
- [59] Arunesh Mathur, Jonathan Mayer, and Mihir Kshirsagar. 2021. What Makes a Dark Pattern ... Dark? Design Attributes, Normative Considerations, and Measurement Methods. In *CHI'21*. ACM, New York, NY, USA, Yokohama, 18. <https://doi.org/10.1145/3411764.3445610>
- [60] Christopher McCrudden and Jeff King. 2015. The dark side of nudging: the ethics, political economy, and law of libertarian paternalism. *Choice Architecture in Democracies, Exploring the Legitimacy of Nudging* (Oxford/Baden-Baden: Hart and Nomos, 2015), Forthcoming, *U of Michigan Public Law Research Paper* 485 (Nov. 2015). <https://ssrn.com/abstract=2685832>
- [61] Thomas Mejttoft, Sarah Hale, and Ulrik Söderström. 2019. Design Friction. In *Proceedings of the 31st European Conference on Cognitive Ergonomics*. ACM, BELFAST United Kingdom, 41–44. <https://doi.org/10.1145/3335082.3335106>
- [62] Thomas Mildner, Orla Cooney, Anna-Maria Meck, Marion Bartl, Gian-Luca Savino, Philip R Doyle, Diego Garaialde, Leigh Clark, John Sloan, Nina Wenig, Rainer Malaka, and Jasmin Niess. 2024. Listening to the Voices: Describing Ethical Caveats of Conversational User Interfaces According to Experts and Frequent Users. In *roceedings of the CHI Conference on Human Factors in Computing Systems* (CHI '24). ACM, New York, NY, USA, Honolulu, HI, USA, 1–18. <https://doi.org/10.1145/3613904.3642542>
- [63] Thomas Mildner, Merle Freye, Gian-Luca Savino, Philip R. Doyle, Benjamin R. Cowan, and Rainer Malaka. 2023. Defending Against the Dark Arts: Recognising Dark Patterns in Social Media. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. ACM, Pittsburgh PA USA, 2362–2374. <https://doi.org/10.1145/3563657.3595964>
- [64] 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*. ACM, Hamburg Germany, 1–15. <https://doi.org/10.1145/3544548.3580695>
- [65] Ministry of Consumer Affairs, Food & Public Distribution. 2023. Central Consumer Protection Authority issues 'Guidelines for Prevention and Regulation of Dark Patterns, 2023' for prevention and regulation of dark patterns listing 13 specified dark patterns. <https://pib.gov.in/PressReleaseFramePage.aspx?PRID=1983994> Accessed 21.12.2023.
- [66] Alberto Monge Roffarello, Kai Lukoff, and Luigi De Russis. 2023. Defining and Identifying Attention Capture Deceptive Designs in Digital Interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–19. <https://doi.org/10.1145/3544548.3580729>
- [67] Donald A. Norman. 2013. *The design of everyday things* (revised and expanded edition ed.). Basic Books, New York, New York.
- [68] Ikechukwu Obi, Colin M. Gray, Shruthi Sai Chivukula, Ja-Nae Duane, Janna Johns, Matthew Will, Ziqing Li, and Thomas Carlock. 2022. Let's Talk About Socio-Technical Angst: Tracing the History and Evolution of Dark Patterns on Twitter from 2010-2021. <http://arxiv.org/abs/2207.10563> arXiv:2207.10563 [cs].
- [69] European Parliament. 2022. Digital Services Act***I. European Parliament [A9-0356/2021]. https://www.europarl.europa.eu/doceo/document/TA-9-2022-0014_EN.html
- [70] Douglas Glen Rizzo, Mario J. Whitman. 2009. Little Brother is Watching You: New Paternalism on the Slippery Slopes Themed Issue: Perspectives on the New Regulatory Era. *Arizona Law Review* 51 (2009), 685.
- [71] 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>
- [72] Hanna Schneider, Malin Eiband, Daniel Ullrich, and Andreas Butz. 2018. Empowerment in HCI - A Survey and Framework. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–14. <https://doi.org/10.1145/3173574.3173818>
- [73] Cass R. Sunstein. 2015. Nudges, Agency, and Abstraction: A Reply to Critics. *Review of Philosophy and Psychology* 6, 3 (Sept. 2015), 511–529. <https://doi.org/10.1007/s13164-015-0266-z>
- [74] Nada Terzimehić, Sarah Aragon-Hahner, and Heinrich Hussmann. 2023. The Tale of a Complicated Relationship: Insights from Users' Love/Breakup Letters to Their Smartphones before and during the COVID-19 Pandemic. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 1, Article 28 (mar 2023), 34 pages. <https://doi.org/10.1145/3580792>
- [75] Richard H. Thaler. 2018. Nudge, not sludge. *Science* 361, 6401 (Aug. 2018), 431–431. <https://doi.org/10.1126/science.aau9241> Publisher: American Association for the Advancement of Science.
- [76] Richard H. Thaler and Cass R. Sunstein. 2008. *Nudge: improving decisions about health, wealth, and happiness*. Yale University Press, New Haven. OCLC: ocn181517463.
- [77] Georgios Theocharous, Jennifer Healey, Sridhar Mahadevan, and Michele Saad. 2019. Personalizing with Human Cognitive Biases. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*. ACM, Larnaca Cyprus, 13–17. <https://doi.org/10.1145/3314183.3323453>
- [78] Amos Tversky and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. 185 (1974), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
- [79] Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. 2019. (Un)informed Consent: Studying GDPR Consent Notices in the Field. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*. ACM, London United Kingdom, 973–990. <https://doi.org/10.1145/3319535.3354212>
- [80] Peter-Paul Verbeek. 2005. *What things do: philosophical reflections on technology, agency, and design* (1. paperback print ed.). Pennsylvania State Univ. Press, University Park, Pa.
- [81] Peter-Paul Verbeek. 2006. Materializing Morality: Design Ethics and Technological Mediation. *Science, Technology, & Human Values* 31, 3 (May 2006), 361–380. <https://doi.org/10.1177/0162243905285847> Publisher: SAGE Publications Inc.
- [82] Ari Ezra Waldman. 2020. Cognitive biases, dark patterns, and the 'privacy paradox'. *Current Opinion in Psychology* 31 (Feb. 2020), 105–109. <https://doi.org/10.1016/j.copsyc.2019.08.025>
- [83] 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>
- [84] 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. In *Proceedings of the 22nd International Conference on World Wide Web* (Rio de Janeiro, Brazil) (WWW '13 Companion). Association for Computing Machinery, New York, NY, USA, 763–770. <https://doi.org/10.1145/2487788.2488038>

- [85] 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. *SOUPS 2011 - Proceedings of the 7th Symposium on Usable Privacy and Security* (2011). <https://doi.org/10.1145/2078827.2078841>
- [86] Christopher D. Wickens, John Lee, Yili D. Liu, and Sallie Gordon-Becker. 2003. *Introduction to Human Factors Engineering (2nd Edition)*. Prentice-Hall, Inc., USA.
- [87] José P Zagal, Staffan Björk, and Chris Lewis. 2013. Dark Patterns in the Design of Games. *Foundations of Digital Games 2013* (2013).