



PhD-FHSE-2024-043
The Faculty of Humanities, Education and Social Sciences

DISSERTATION

Defence held on 19/11/2024 in Esch-sur-Alzette

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN *PSYCHOLOGIE*

by

Lorena SÁNCHEZ CHAMORRO

Born on 30th October 1993 in city of Madrid, (Spain)

*DISENTANGLING VULNERABILITY TO MANIPULATIVE
DESIGNS: AN EXPERIENTIAL PERSPECTIVE TO RETHINK
RESISTANCE STRATEGIES*

Dissertation defence committee

Dr. Carine Lallemand, dissertation supervisor
Assistant Professor, Université du Luxembourg

Dr. Colin M. Gray, Vice Chairman
Associate Professor, Indiana University Bloomington

Dr. Luis Leiva, Chairman
Assistant Professor, Université du Luxembourg

Dr. Hanna Schraffenberger
Assistant Professor, Radboud Universiteit

Dr. Daniel Fitton
Associate Professor, University of Central Lancashire

Lorena Sánchez Chamorro

Disentangling vulnerability to manipulative designs

An experiential perspective to rethink
resistance strategies



Disentangling vulnerability to manipulative designs

An experiential perspective to rethink resistance strategies

Doctoral dissertation by Lorena Sánchez Chamorro

University of Luxembourg

ACKNOWLEDGEMENTS

During one of my readings to prepare this dissertation this summer, I discovered a new passion: reading the preface of scientific books. In one of them, I found that the Nobel Prize Daniel Kahneman felt that his habitual colleague and collaborator, Amos Tversky, always complemented (and improved) his ideas during their conversations. I also found that Heidegger wrote his most influential work, “Being and Time”, in a rush because he needed to secure his tenure track position. Consequently, he felt the need to write a second part that compensated for the first “unpolished” one. That second part never saw the light.

In a summer school about “Ethics in design”, I learnt that to understand your core values deeply, one has to listen carefully to the “light-bulb” moments: those moments in which something in your guts calls you, something moves you or feels off — or right. Those are the moments worthy of exploration and introspection. While writing this manuscript, I have often thought about those pieces of intra-history of science, not only because I found them quite surprising but also because they came to me as “light-bulb” moments that have given me so much perspective. The perspective of my process during these years, but also the perspective of what is yet to come: who I want to be as a researcher and the role of academic research to change the world. Indeed, I started my PhD telling all my potential supervisors “*I want to do a PhD because I want to change the world*”. I know it sounded naïve, but that core value has not changed ever since.

Of course, those “light-bulb” moments would have never come without my supervision committee’s invaluable support and guidance. I am immensely grateful to have received constant inspiration, critical thinking, and trust. **Gabriele Lenzini**, thank you so much for your transparency, honesty, and for “practising what you preach”. Thank you for standing up for your values and even for fighting for them when you consider it necessary. We need more scholars like you in academia. **Colin Gray**, I am immensely thankful for your guidance and support. Thanks for getting me out of my comfort zone with every study and every question. I am incredibly grateful for your openness, kindness and constant willingness to inspire and teach others — even when those others introduced themselves to you with the worst of introductions.

The deepest of my gratitude goes to **Carine Lallemand**. I have no words to express appreciation for your immense support. Thank you for your trust, for believing in me when nobody else did — not even myself —, and for your constant reminders to believe in myself more often. As a great philosopher once said, “must be exhausting always rooting for the anti-hero”. Thanks for being an extraordinary mentor and role model; for being fierce and brave. You are the “safety net” every doctoral candidate wishes to have.

I am deeply honoured to have made this journey with marvellous, brilliant people. **Romain Toebosch**, our lengthy and intense discussions have not only contributed enormously to my work and how I see the field but also to who I want to be as a researcher. You taught me that both design and research are a continuum. Thanks for supporting me, challenging me, and inspiring me every day.

I also want to express my deepest gratitude to the organisations and people that made

my data collection possible: **Asociación La Rueca, La Golondrina, FAMASA, Asociaciones de Vecinos de Tetuan y Amistad de Canillejas, and Unión Democrática de Pensionistas y Jubilados de España**. Thank you so much **Sara Cañedo, Maite Cuenca, Carlos Martínez, Paco Gómez, Francisco Javier Vicente, and Carmen Page** for trusting me, and opening the doors of your organisations with generosity. Thank you, **Belén Gilaranz and Julián Vicente**, for always saying “yes” to every crazy idea I propose. Above all, **I will always be in debt to the participants of my studies** who generously shared their experiences with me and helped researchers to feel connected to reality: they are the reason why I do this.

I have been lucky to work on a pioneering, exciting, multidisciplinary, and full-talent project. **Arianna Rossi and Anastasia Sergeeva**, thanks for being my companions against manipulative designs. Even if the battle was long and chaotic, it was definitely worth it. I must, therefore, thank my fellow members of the DECEPTICON project, especially **Vincent Koenig and Kerstin Bongard-Blanchy**, who gave me the opportunity to conduct this amazing research.

I would love to mention my gratitude to the manipulative design research scholars who have accompanied me during this dissertation, not only with their amazing papers, but with their engaging discussions, debates, and challenges: **Cristiana Santos, Thomas Mildner, Johanna Gunawan, Hauke Sandhaus, Hanna Schraffenberger, Frederik Zuiderveen, Ritika Gairola, Silvia de Conca, Raphael Gellert, Katie Seaborn, Nataliia Bielova, René Schäfer**, and all the brilliant minds that gathered in the Lorentz Workshop.

Among the many things I have been doing these years, I was lucky to advise and work with **Paulo, Adel, Iago, and Clément** to design a fairer video-game ecosystem. Thank you, guys, for bringing me so much inspiration.

I would also like to thank the members of the HCI team — and beyond — who have supported me along the way: **Luce, Verena, Sophie, Xiaowei, Hailey, Vincent, Mateusz, Luis, Isabelle and Bereket**. I would also love to mention my special thanks to **Claudia Negri-Ribalta** for being such an amazing inspiration. **Alina**, thanks for making Luxembourg a brighter place. **Margault**, thanks for making Luxembourg the place I am going to miss.

This adventure would not have been possible without **Juanjo**, who was the first to see the researcher in me and brought me here. Words cannot express my gratitude for what you brought to my life.

I could not have survived the cold days of Luxembourg without **Ichi and Backe**. Thank you for making me feel so close everyday, while being so far apart.

Thank you also, **Nathalie, Oscar, and Arthur**, for suffering this thesis with me, and for your extensive and unconditional support in the middle of the chaos.

I am deeply indebted to my parents, **Yolanda and Paco**. Because of them, I am an insatiable learner, a fierce fighter, and a passionate researcher. Thank you for being so proud of every little achievement I may have. Thank you for having me on your minds, and on your conversations. **Andrés**, I feel immensely honoured to have you; thank you for all the night conversations that inevitably ended up in Carlos Marx. Because this thesis also has a part of you. **Romain**, thanks for not letting me fall, for sharing with me your passion, your ideas, and the biggest of the smiles every time that Luxembourg stood between us.

A mi abuela Rosa. Porque esto es lo que estaba haciendo cada vez que me preguntabas
"cuánto tiempo me quedaba para volver de 'mi pueblo'".

Thesis Summary

Manipulative designs or so-called “dark patterns”¹ are design features, patterns and mechanisms that “subvert, impair, or distort the ability of a user to make autonomous and informed choices in relation to digital systems regardless of the designer’s intent” [174]. By steering users to make decisions they would not make if fully informed, manipulative designs put users’ autonomy at stake, which is associated with a wide range of harms. Some people suffer the consequences of manipulative designs more than others: they are more vulnerable to the harms of these designs. Vulnerability is a position of power imbalance in which users are more susceptible to receiving an impact and less likely to recover from it. Vulnerability is layered, situated, and interpersonal: while everyone can be vulnerable, some can be more vulnerable than others, and the “drivers of vulnerability” are multifaceted elements that place users in such positions of higher risk. With this in mind, I investigated how HCI and design scholarship can contribute to rethinking countermeasures to protect users by understanding the experiences that make users vulnerable to manipulative designs. Therefore, the overall objective of this dissertation is to understand what vulnerability means in the realm of manipulative designs and to help the design community integrate this knowledge into theory and practice.

This dissertation first examines practitioners’ perspectives and explores how the tensions between persuasive design and manipulation in UX design practice inform vulnerability to manipulative designs. It explains how manipulation can be engineered in interaction design and what tensions practitioners face in their design processes by investigating experienced UX/UI designers with co-creation workshops. This study also provides design guidelines to support practitioners. Following that study, this dissertation examines how HCI can contribute to conceptualising vulnerability in manipulative designs through a multidisciplinary conversation. By understanding the flaws of legal texts in conceptualising vulnerability, we present the different ways in which users may become vulnerable, with a special focus on their ecologies, and provide some tools for legal scholars and policymakers to learn from HCI and design expertise.

Building on the idea that context can make users vulnerable, this dissertation explores the contextual aspects that drive vulnerability to manipulative interfaces. To do so, I then report on three main studies relying on qualitative and design research-inspired methods with three traditionally considered vulnerable groups —teenagers, young adults at social exclusion risk, and older adults. First, by studying teenagers’ experiences in three contexts – video games, social media and e-commerce –, this thesis explains the social aspect of manipulative designs and provides contextual harms tied to this population. With this study, I highlight the importance of social relationships as mediators of experiences with manipulative designs. Second, with “magic machines workshops” to understand the experiences of manipulation in older adults, the thesis showcases their needs regarding resisting manipulation. How older adults understand manipulation helped us identify design challenges for counter-interventions. Third, by understanding young adults with lower levels of digital skills and their

¹The research community is studying this phenomenon using a variety of labels, including deceptive design, nudges, anti-patterns, and most dominantly, “dark patterns.” Following the ACM recommendations on diversity and inclusion [143] I hereby use the term “manipulative designs” to describe this phenomenon.

experiences with manipulative designs, this thesis explains how the imaginaries of manipulative designs are related to the different ways to resist them. Lastly, through a scoping review of existing intervention spaces and the development of an experience map, I discuss potential intervention spaces and design challenges for the community that aim to target situations of vulnerability.

Overall, this thesis contributes to the empirical investigation of vulnerability to manipulative designs, reflecting on the experiential vulnerability drivers. Social drivers, interaction drivers, and drivers related to users' agency mediate their interactions and make them vulnerable. By studying what vulnerability means in the context of manipulative designs, I also identified challenges and opportunity spaces to design counter-interventions for manipulative designs and to prevent users from experiencing situations of vulnerability. This has allowed me to suggest a future direction for the community that reframes the problem of online manipulation as one of vulnerability. To help scholars in this, I provide a template to start with new problems, nuances and approaches for studying and preventing manipulative designs.

This research has been possible thanks to the Fonds National de la Recherche under the grant no. IS/14717072 Deceptive Patterns Online (Decepticon).

Table of Contents

I Introduction and Motivation	3
1 Introduction	5
1.1 The Omnipresent Problem of Online Manipulative Designs	7
1.2 Vulnerability and the Experience of Manipulative Designs	8
2 Research Approach	11
2.1 Situating the Problem of Vulnerability to Manipulative Designs	13
2.2 Theoretical Approach	16
2.3 Objectives and Research Questions	18
2.4 Thesis Overview	21
3 Related Work	25
3.1 Introduction	27
3.2 Method	27
3.3 Findings	29
3.4 Provocations	33
3.5 Experiences with Manipulative Designs	35
4 Methodology	43
4.1 Selection of Methods	45
4.2 Selection of Research Methods	48
4.3 Collaboration with Study Participants	49
4.4 Data Analysis	50
4.5 Positionality Statement	51
II Conceptualising Manipulation in Practice	55
5 Ethical Tensions in UX Practice	57
5.1 Introduction	59
5.2 Related Work	60
5.3 Methodology	64

5.4	Findings	68
5.5	Discussion	76
5.6	Design recommendations for practitioners	80
5.7	Limitations and Future Work	82
5.8	Conclusion	82
5.9	Chapter Takeaways	83
6	Who is Vulnerable	85
6.1	Introduction	87
6.2	Current conceptualisation of vulnerability	89
6.3	Reframing Vulnerability in Digital Markets	91
6.4	Risk of Vulnerability and Harms	93
6.5	Risk Assessment and Vulnerability	95
6.6	Factors Influencing the Risk of Vulnerability	103
6.7	discussion	113
6.8	Chapter Takeways	117
III	Contextual Drivers of Vulnerability	119
7	Teenagers' Experiences with Manipulative Designs	121
7.1	Introduction	123
7.2	Related Work	124
7.3	Methodology	126
7.4	Findings	130
7.5	Discussion	139
7.6	Chapter Takeaways	143
8	The Experiences of Manipulation in Older Adults	145
8.1	Introduction	147
8.2	Related Work	148
8.3	Methodology	151
8.4	Findings	157
8.5	Discussion	166
8.6	Conclusion	171
8.7	Chapter Takeaways	171
9	Manipulation and Resistance	173
9.1	Introduction	175
9.2	Related Work	176
9.3	Methodology	179
9.4	Findings	186
9.5	Discussion	198
9.6	Conclusion	202
9.7	Chapter Takeaways	203

IV Design Spaces to Mitigate Vulnerability to Manipulative Designs 205

10 Mapping the Experience of Online Manipulation	207
10.1 Introduction	209
10.2 Related Work	210
10.3 Research Approach	214
10.4 Study 1. Systematic Review of Interventions	215
10.5 Study 2. Mapping the Experience With Manipulative Designs	231
10.6 Discussion	242
10.7 Conclusion	250
10.8 Chapter Takeaways	250
11 The Disentanglement	253
11.1 The Entangled Problem of Manipulative Designs	255
11.2 The Disentanglement	256
11.3 The Disentanglement for HCI	261
11.4 The Disentanglement for Practitioners	269
11.5 The Disentanglement for Policymakers	271
11.6 Users as Stakeholders: A Critical Perspective on Educational Interventions . . .	272
11.7 Overall Limitations	273
12 Concluding Remarks	275
13 Appendix	309
13.1 Chapter 5	309
13.2 Chapter 8	310
13.3 Chapter 9	312

GLOSSARY OF ACRONYMS

ACM — Association for Computing Machinery

ACM DL — Association for Computing Machinery Digital Library

AI — Artificial Intelligence

CSCW — Computer-Supported Collaborative Work

DSA — Digital Services Act

DPIA — Data Protection Impact Assessment

DPO — Data Protection Officer

EC — European Commission

EDPB — European Data Protection Board

FF — Fast Fashion

FOMO — Fear of Missing Out

GDPR — General Data Protection Regulation

HCI — Human-Computer Interaction

NGO — Non-Governmental Organisation

OECD — Organisation for Economic Co-operation and Development

STS — Science and Technology Studies

UCPD — Unfair Commercial Practices Directive

UI — User Interface

UX — User Experience

Part I

Introduction and Motivation



This part establishes the motivations for this work. **Chapter 1** introduces the problem of manipulative designs and its relationship with vulnerability, by establishing the directions of the dissertation.

Chapter 2 explains the research approach and explains the importance of understanding users' lived experiences to better comprehend the problem of online manipulative designs. It also sets the objectives and research questions.

Chapter 3 provides a critical review of the literature on manipulative designs and **Chapter 4** gives the overview of the methodology defined to investigate vulnerability in the realm of manipulative designs

Chapter 1

Introduction

“The unprecedent is necessarily unrecognizable. When we encounter something unprecedented, we automatically interpret it through the lenses of familiar categories, thereby rendering invisible precisely that which is unprecedent.”

Shoshana Zuboff, *The Age of Surveillance Capitalism*, 2019

1.1 The Omnipresent Problem of Online Manipulative Designs

How many times have you accepted a cookie consent banner because it did not have a visible reject option? Or how many times have you booked the slightly more expensive hotel because there was “only one left” and you did not want to risk losing it for the opportunity of finding something better? These are not singular, coincidental experiences, but common events triggered by manipulative designs.

Manipulative designs, or so-called “dark patterns”¹ are design features, patterns and mechanisms that “subvert, impair, or distort the ability of a user to make autonomous and informed choices in relation to digital systems regardless of the designer’s intent” [174]. Alterations in the online choice architecture leading users make decisions against their own interests [297, 308], nagging pop-ups asking users to make the same decisions several times [315, 174], and cues that reflect the scarcity of a product to induce impulsiveness in users [439, 396] are some common examples of these manipulative practices. These design instances flourished with the enhancement of user experience (“UX”) techniques and the institutionalisation of UX as an important component in the development of websites and platforms [133, 322]. Manipulative designs, hence, arise as a distortion of UX practices, blurring a line between what is ethical or not in UX design [169]. This blurred line causes confusion among designers about what is permissible or not, and while partially these designs may come from well-intended “bad design” [86], the use of these strategies has been extended among all kinds of services — e.g. newspapers [401], e-commerce sites [228, 328, 447], or social media [390, 232, 309].

The problem that Harry Brignull pointed out in 2010 — coining the term “dark patterns” [55] — crystallised within the research community in 2018 with the pioneering work of Gray et al. [169], which explored the “dark side” of UX design practices. Since then, attention to these designs has increased at the same pace as their proliferation in all kinds of sectors and digital services — social networks, privacy consent banners, travel sites, streaming platforms, and e-commerce. By steering users into decisions they would not make if fully informed, manipulative designs put users’ autonomy at stake, which is associated with a wide range of cognitive, psychological, financial, well-being, or privacy-related harms, among others [181, 386]. Their relationship with users’ harm has increased the attention of scholars and policymakers who are trying to define solutions to overcome these negative effects. The Organisation for Economic Co-operation and Development (“OECD”) released a comprehensive report about the harms and potential regulation of these practices [341]. In Europe, the Digital Services Act [354] has included the term “dark patterns” in its Recital 67 as an antecedent to explain that *“providers of online platforms shall not design, organise or operate their online interfaces in a way that deceives or manipulates the recipients of their service or in a way that otherwise materially distorts or impairs the ability of the recipients of their service to make free and informed decisions*

¹The research community is studying this phenomenon using a variety of labels, including deceptive design, nudges, anti-patterns, and most dominantly, “dark patterns.” Following the ACM recommendations on diversity and inclusion [143] I hereby use the term “manipulative designs” to describe this phenomenon.

practices" in Article 25. Different authorities in data protection or consumer protection have released several guidelines to address the problem in digital services [343, 41]. Outside Europe, several countries are trying to specifically address these practices for a longer time in the United States [56, 270], and with more recent incorporation into the regulatory landscape in India [400]. Yet, several challenges remain in the design practice to release websites from these manipulative instances. To achieve such an endeavour, the first step is questioning when users become vulnerable to these manipulative practices.

Vulnerability is understood as a position of power imbalances in which users are more susceptible to receive an impact and less likely to recover from it [288], and feeds the discussion of manipulative designs in two directions. On the one hand, the extent to which specific groups traditionally labelled as vulnerable [138, 283, 340, 341, 25, 300] — e.g. kids, older adults, or minorities — can be more impacted by these design features is in the public and scholarly debate [41, 340]. On the other hand, the fact that vulnerability to design features has been potentiated by surveillance capitalism and the data economy makes all users vulnerable [194, 288] highlighting the need to study the specific design features that might increase the vulnerability of users. To do so, this dissertation draws on current discussions about vulnerability and understands vulnerability as layered, situated and interpersonal [138, 283, 288]: while everyone can be vulnerable, some can be more vulnerable than others.

Since recognising manipulative designs is a hard task for users [286, 45, 309, 111], scholars and policymakers try to protect users from harms instigated by these designs. This is not an easy task, as the line between persuasive and manipulative designs is not always evident — the former being traditionally accepted in the design community [141, 338, 142]. It is not the intention of this dissertation to justify — or criticise — persuasive design but to invite the community to reflect on the conditions under which persuasion crosses this line into manipulation. Thus, the ethical limits of persuasive design are intimately related to the permissibility of manipulative designs. Following Susser et al. [429] problematisation of online manipulation, persuasion is self-evident for the user that is being persuaded, while manipulation tries to steer users in a hidden way "by exploiting their vulnerabilities" [429].

The problem with its definition poses the need to understand what makes users more vulnerable to manipulative designs from different angles. If *HCI and design researchers* want to empower and protect users via design artefacts, we need to understand the factors that make users vulnerable. If *design practitioners* want to use persuasive design techniques, they need to be aware of vulnerability factors in their design processes to evaluate their designs and avoid manipulation. If *policymakers* aim to ban technologies and designs that manipulate users and may impose risks to autonomy, privacy or well-being, they need to know what makes users vulnerable and prone to those risks.

1.2 Vulnerability and the Experience of Manipulative Designs

Understanding what makes users vulnerable to manipulative designs becomes a crucial issue for all stakeholders that aim to protect users. Scholars from different domains approach

the problem in their own ways. The HCI and design community is examining the problem from an interaction design point of view, understanding that the different design elements have the power to change behaviours in users (e.g. [31, 38, 336]). For instance, colours, or elements' position might determine the behaviour of the users when it comes to interacting with a manipulative design (e.g. [183, 31, 46, 247]). This approach, that often corresponds to second wave of HCI traditions [149, 191], overlooks the importance of the contextuality of the problem.

Scholarship on manipulative designs has increased [171, 76] and investigated their presence in different domains, like cookie consent banners [31, 46, 183, 38], social media [309, 308], video platforms [281, 78], travel services, e-commerce [447, 439], robots [399], newspapers [401], augmented reality [128, 456] and digital services in general [45, 286]. Yet, research that looked at socio-demographic mediators like age [310, 45, 280, 15], or education [45, 280] is limited; as are the works that include diverse populations [393]. Evaluative approaches that look at the existence of manipulative design [231, 247, 234], and experimental settings that try to see changes in behaviour in the presence of manipulative designs are popular among scholars [447, 38, 336]. However, there is still a gap in understanding the contexts of non-normative users and how different socio-demographic factors and specific experiences make users interact with manipulative designs in the way they do.

Experiences with manipulative designs are shaped by the different ecologies in which the interaction takes place. Ecologies, understood as nested systems that mediate the interaction — macro, meso, and microsystems — [57, 320, 220, 97] play a role in how users interact with manipulative designs, and therefore contain drivers of potential vulnerability [288]. This dissertation argues that manipulative and deceptive designs occur when one or more factors give rise to specific contextual and situated experiences that make users vulnerable, which will be unveiled along the length of this manuscript.

The search for solutions and interventions also cannot overlook that user interactions are situated and contextual; hence, the socio-technical context of users not only matters to understand their relationship with technology in general, but with manipulative designs in particular. This contextual, interpersonal element complicates the study of vulnerability towards manipulative designs and fuels new design challenges when it comes to the definition of countermeasures. Thus, while different interventions are being proposed to protect users — solutions such as friction [45, 316], preventing Fear-of-Missing-Out ("FOMO") design [458], or goal-setting features [281, 285] —, it is necessary to question whether those are equally effective in helping users in situations of increased vulnerability, prompting us to look at interventions through the lens of vulnerability for a more nuanced perspective of the problem and potential interventions.

In this dissertation, I reject the idea of the homogeneous user and look at the experience of manipulation in diverse populations in situations of vulnerability to understand the drivers that contribute to their experience. While I embrace and agree with the critics of reducing vulnerability to a label, which can be stigmatising [138, 283], I look at the experience of populations traditionally considered vulnerable in order to understand their contexts and

ecologies, eliciting their specific drivers of vulnerability, ultimately detaching this work from those labels. In this way, the ecologies, contexts and experiences analysed in this dissertation are not exclusive to the people that fall under the original labels associated with the studied populations: the drivers and solutions can be extrapolated to others that share similar experiences. Thus, the design challenges discussed in this dissertation are not tied to one specific population — e.g. older adults or teenagers — but are useful to define interventions for all kinds of people that share the same drivers of vulnerability.

The overall objective of this dissertation is, therefore, to understand what vulnerability means in the realm of manipulative designs through a nuanced understanding of the experiences with manipulative designs, and to help the design community to integrate that knowledge into design theory and practice. In the following chapter, I will operationalise this objective into the research questions this dissertation addresses and present the theoretical and epistemological approaches underpinning this work.

Research Approach

Abstract. This chapter introduces the main theoretical lenses around which this dissertation is framed and explains how the problem of vulnerability to manipulative designs is conceptualized. First, I will introduce “vulnerability” to manipulative designs as a problem that needs to be understood in HCI and design scholarship. Second, I will describe the main schools of thought used to drive the objectives, and research questions of this dissertation.

2.1 Situating the Problem of Vulnerability to Manipulative Designs

2.1.1 Manipulative Designs as a Design Problem

Before discussing the relationship between vulnerability and manipulative designs, it is important to identify the relevance of manipulative design practices for interaction design as a field. The idea of “nudge” as an alteration of the choice architecture to steer certain behaviours on citizens has been discussed in the realm of public policy and economics, behind the idea of soft paternalism — i.e. altering behaviours without changing economic incentives for the citizens’ own good [431]. Under such rationale, some advocates of nudge theory consider “nudges” as ethical, while other scholars have determined specific conditions of transparency for those to be ethical [304, 233, 330].

Some authors have considered “sludge” as the opposite of nudges, and concretely “digital sludges” to name online manipulative practices. However, manipulative design transcends such discussion in two ways: (i) the alterations are made against the users’ best interests, attacking their autonomy by exploiting their vulnerabilities, and (ii) the interface becomes the entry point for different harms. Although this dissertation does not aim to debate nudge theory in the online domain, Chapter 5 deepens into the position of manipulative designs as instances that coerce, manipulate, and deceive users, with relevance in interaction design. Similarly, marketing and business studies have investigated some types of manipulative designs with the purpose of making more effective strategies rather than protect users (e.g. [27]), hence that set of literature is out of the scope of this dissertation.

This dissertation uses the nomenclature and tradition of scholars in HCI such as Mathur et al. [297] and Gray et al. [174], whose work strives towards developing more ethical user interfaces acknowledging the potential harm of manipulative designs. I, therefore, understand manipulative designs as mechanisms that “subvert, impair, or distort the ability of a user to make autonomous and informed choices in relation to digital systems regardless of the designer’s intent” and use the main 6 categories of high-level patterns specified by Gray et al. [174]¹. I chose the terms provided in Gray et al.’s ontology given it is the most comprehensive glossary of terms and definitions existing in the community since they took the main existing taxonomies in scholarship and policymaking and systematised them into a final comprehensive ontology.

1. Sneaking “hides, disguises, or delays the disclosure of important information that, if made available to users, would cause a user to unintentionally take action they would likely object to” [174], a low-level example of this strategy is “disguised ads” that use style elements that imitate news or other websites and that are not marked as an advertisement.

¹Some of the studies included in this dissertation were conceptualised before the final ontology was developed. Therefore, I considered the initial 5 strategies from Gray et al. [169]

2. Obstruction “impedes a user’s task flow, making an interaction more difficult than it inherently needs to be, dissuading a user from taking action,” [174] like creating barriers for users.
3. Interface interference “privileges specific actions over others through manipulation of the user interface, thereby confusing the user or limiting discoverability of relevant action possibilities” [174] like the manipulation of the hierarchy within the choice architecture.
4. Forced action — “requires users to knowingly or unknowingly perform an additional and/or tangential action or information to access (or continue to access) specific functionality, preventing them from continuing their interaction with a system without performing that action” [174], an example is nagging techniques that ask users for a task they have already completed.
5. Social engineering — “presents options or information that causes a user to be more likely to perform a specific action based on their individual and/or social cognitive biases, thereby leveraging a user’s desire to follow expected or imposed social norms” [174] like in the case of playing with urgency or scarcity messages.

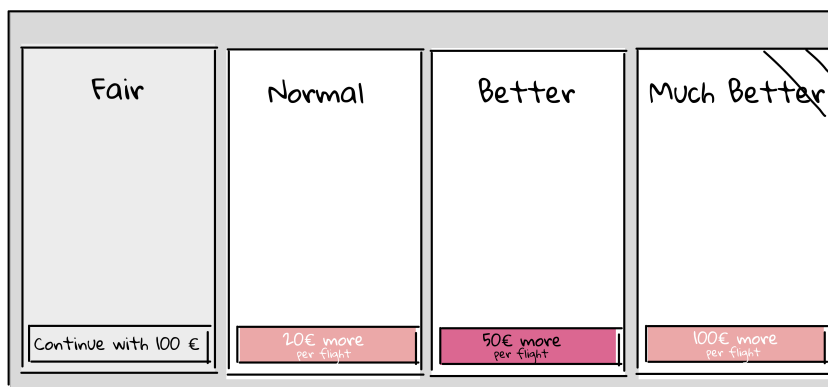


Figure 2.1: Sketch of an interface with manipulation of choice architecture in which colours aim to induce users to select a more expensive one.

2.1.2 Understanding the Debates of Vulnerability Online

Vulnerability is a contested concept and a broad multifaceted construct. In general terms, it refers to the users’ position of higher exposure to receive and impact and to be less likely to recover from them [290]. However, their lack of concreteness poses challenges for policymakers to regulate and scholars to design accounting for vulnerability.

In the design and interaction with technology, Barta et al. [25] explain how HCI and CSCW scholars have understood vulnerability as harm — some groups would be more harmed than

others — and as “risk and reward” (p.6) — disclosing can make you vulnerable, especially in some contexts like social media. From the HCI perspective, platforms are seen as contributors to such socio-technical harm [392]. Yet, the nature of HCI as a discipline makes scholars often use *vulnerability* to understand the experiences of certain groups, as a proxy, without the intention of questioning its source.

The source of vulnerability, and how to respond to it has been a part of the conversation in legal theory. To briefly situate the debate of vulnerability, and by extension, online vulnerability, scholars in legal theory question two approaches: (1) the personalistic approach where specific users are deemed vulnerable — e.g., children, older adults, minorities — [283], and (2) the universalistic approach where everyone can be vulnerable because vulnerability is inherent to the human condition [138].

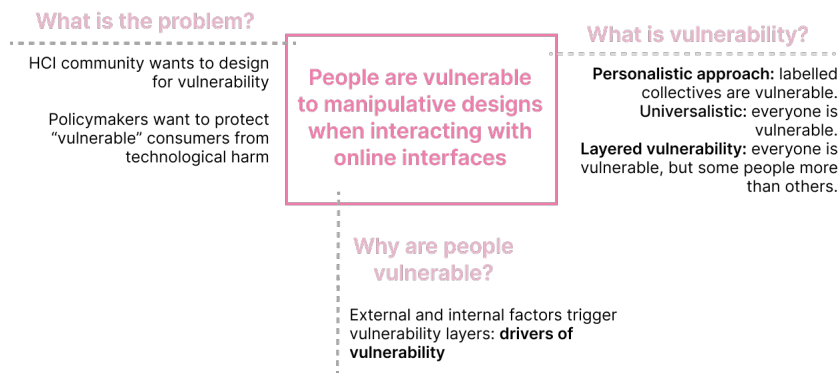


Figure 2.2: Representation of the relationship between manipulative designs and vulnerability

2.1.3 Layered Vulnerability to Manipulative Designs: When Context Matters

Reconciling legal and HCI perspectives, when it comes to vulnerability, will allow the use of the tools that HCI provides to empirically investigate vulnerability and to translate it in theoretical terms that policymaking requires [173]. Acknowledging vulnerability to manipulative designs as a platform-enabled type of vulnerability [194, 392], I build on theoretical approaches of vulnerability. Within this debate, I follow the lead of Malgieri [288], who builds on Luna’s theories of layers of vulnerability [283]: everyone is vulnerable, but some are more vulnerable than others given the existence of *layers of vulnerability*. Layers of vulnerability, as opposed to categories, are different user conditions — e.g. personal, contextual, physical — that accumulate and overlap, providing a continuum of vulnerability rather than a category in which users belong or not.

According to Luna [283], layers of vulnerability need to be triggered to unchain vulnerability; these triggers are what Malgieri understands as “drivers” [288]. For Malgieri, vulnerability relates to power imbalance, and there are multi-faceted elements that mediate the relation-

ship between users and technological providers that fuel those power imbalances, putting users at a higher risk: *drivers of vulnerability*.

While Luna and Malgieri remain at a theoretical level, by explaining how some factors — e.g., age, gender, race — contribute to these layers and drivers of vulnerability, I aim to explain how those layers relate to experiential aspects of the interaction. I, therefore, build on the idea of interaction-as-embodied-action [207], in which technology is not a mere tool but a part of the user. As it will be further detailed in the following sections, the way users interact with technology is conditioned by the contexts in which users are and their situated experiences.

The OECD, in their report on digital vulnerability, also builds on the theories of vulnerability and provides some examples of drivers that belong to personal characteristics, behavioural factors, and markets [340]; yet, they overlook the contextual and situated aspect of vulnerability. Thus, beyond the factors pointed out by personalistic views of vulnerability, context can drive vulnerability within the interaction with manipulative designs. This dissertation draws on the idea that, on specific contexts, manipulative designs act by triggering layers of vulnerability. It is, hence, the aim of this dissertation to investigate the situated contexts in which the interaction takes place to understand how vulnerability is driven in the presence of manipulative designs.

2.2 Theoretical Approach

To understand what drives vulnerability to manipulative designs it is necessary to look at the lived experiences of users. I here explain the main lenses used for this dissertation. First, I will explain why context matters when it comes to studying vulnerability. Second, how socio-digital inequalities matter, given they impact the context in which the interaction takes place. Third, I will outline how such contextual interactions can only be understood through lived experiences.

Context impacts cognition, perception and interaction with manipulative designs

Ecological theorists illustrate the reasons we behave and how we interpret affordances — as opportunities for behaviour — in the world are determined by perception [160], but the internal process of learning is embedded and intertwined within ‘nested structures’ that surrounds users [57](p.3). The physical, social and material environment in which users live and interact with technology shape them, this is what Bronfenbrenner [57] coined as macro, meso, and micro-systems that contribute to human development. While the functioning of these systems will be further explained in Chapter 6, for now, it suffices to say that macro systems relate to macro-structures in which the user is — e.g. economic system —, meso systems relate to everyday environments of users — e.g. neighbourhood —, while micro systems refer to specific individual users conditions — e.g. users’ age.

Socio-economic conditions impact contexts in which the interaction happens

Socio-digital inequalities are the differences in the acquisition of technology, skills, and outcomes obtained by users from technology. Socio-economic and contextual factors have an impact on those differences [195]. These inequalities are defined as ‘socio-digital’ because socio-economic causes are associated with these differences in the outcomes of technology. With a similar rationale, DiPaola and Calo [113] call for understanding “socio-digital vulnerability”. Resonating with Helsper’s perspective, the socio-economic environment mediates the way users interact with technologies and, in our case, manipulative designs. While social aspects influence the interactions with the online domain, interactions also have social consequences: the interface and its design act as a hinge to the experience of vulnerability in the online domain.

Thus, not paying attention to identities of vulnerability in design might contribute to reinforcing inequalities [300] — e.g. considering norms and standard assumptions in the design of artefacts when some users do not fit such assumptions. As a consequence, the design of artefacts contributes to afford, or, alternatively, disafford interactions [98] and, therefore, allows users — or not — to take benefit from technology. She explains how affordances do not work for everyone in the same way, since some users may be neglected by design, especially for those users that belong to the categories of epistemic injustice along the matrix of domination — e.g., male vs female, cis vs. trans, healthy vs. unhealthy. The idea of layered vulnerability is especially helpful in the context of intersectional conditions [98]: since different conditions may overlap and intersect, looking at situated layers and contexts rather than labels provides a more nuanced view of vulnerability. By adhering to this school of thought, I reject the idea of the homogeneous user: no one-size-fits-all in the realm of manipulative designs.

Accounting on lived experiences to understand drivers of vulnerabilities to manipulative designs

With this dissertation, I aim to articulate how vulnerability is driven in the presence of manipulative designs. To conceptualise vulnerability drivers, caring about the lived experiences of users is fundamental. I here follow the lead of third-wave of HCI scholars that value subjective experiences as mediating points between users and interactions as they are situated, embodied and contextual [191, 123].

Studying the user experience with manipulative becomes indeed fundamental since manipulative designs are sometimes unnoticed by users, as they mean to be subtle [280, 45]. Inspired by the work of Star [424], from Science and Technology Studies, I argue that manipulative designs are relational. In the same way that users only notice a water pipe in relation to their actions — e.g. filling a glass of water, or having to repair the system —; manipulative designs are visible to the user in relation to their own experiences. They are visible to users when they experience an impact caused by their interaction, and some previous works already hinted towards that idea [169, 287]. Gray et al. [169] already hinted at this idea when conceptualising manipulative designs as a mismatch between users’ expectations and the result of

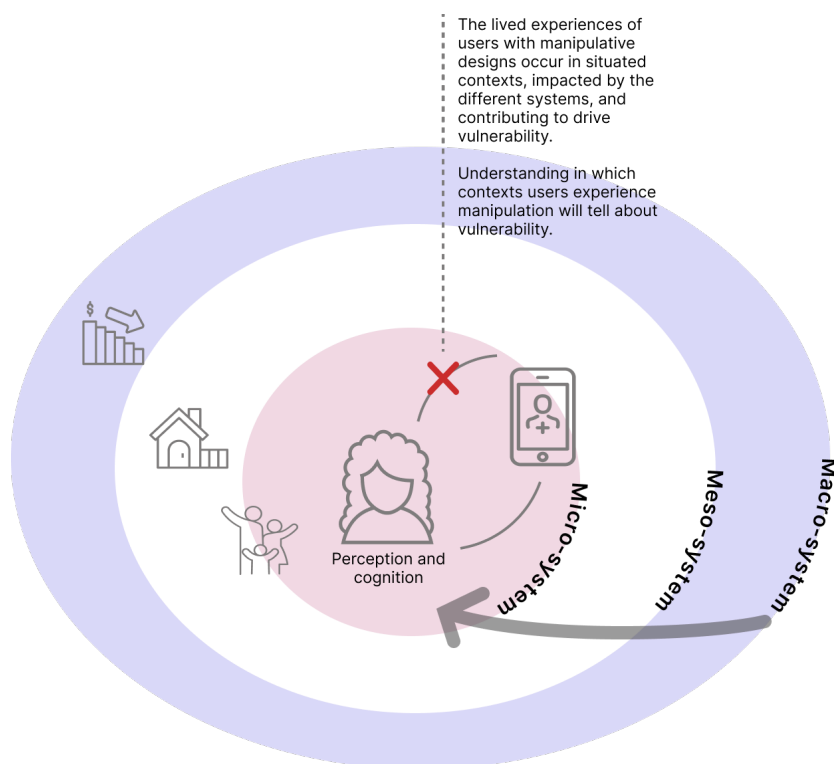


Figure 2.3: Schematic representation of how the different macro, meso, and micro-systems impact users' experiences and interactions with manipulative designs, which lead to perceive manipulative designs and harm differently. Understanding the contexts in which these interactions happen will help to understand users' vulnerability.

the interaction. Similarly, Maier and Harr [287], in an exploratory qualitative study, showed how participants assessed the severity of manipulative designs in relation to their potential impact. Indeed, M. Bhoot et al. [286] found the users' frustration level and the perception of misleading behaviour from the interface to be strong predictors of manipulative design identification, which aligns with the idea of the relationality of these designs.

In light of these aspects, looking at the lived experiences of manipulation will help to understand drivers of vulnerability (See Figure 2.3).

2.3 Objectives and Research Questions

The main objective of this dissertation is to explore the users' drivers of vulnerability in their relationship with manipulative designs. Informed by these insights, the present work additionally aims to explore design solutions that could help users resist — or be resilient to — manipulative designs. This dissertation revolves around the following research questions, which aim to address different aspects of this phenomenon.

Research Question	Focus
RQ1. How do the tensions between persuasive design and manipulation in UX design practice inform vulnerability to manipulative designs?	Conceptualising the problem — Chapter 5
RQ2. How can HCI contribute to the conceptualisation of vulnerability to manipulative designs?	Contextualise the problem — Chapter 6
RQ3. What are the contextual aspects that drive vulnerability to manipulative interfaces?	Exploration of drivers — Chapters 7, 8, and 9
RQ4. What are intervention spaces to design mitigation strategies for manipulative designs that account for vulnerability?	Exploration of the intervention space — Chapter 10

Table 2.1: Research questions, focus and related thesis chapters

RQ1. How do the tensions between persuasive design and manipulation in UX design practice inform vulnerability to manipulative designs? This question helps to contextualise the problem of manipulative designs in practical terms, and why understanding vulnerability is important. By working with practitioners, I aimed to have the point of view and tensions that UX designers experience every day when they design to influence people. I looked at designers' practices through a theoretical grounding. However, these theoretical lenses come from what constitutes manipulation since the problem of manipulative designs is rooted in ill-intended influential practices — deception, coercion and manipulation —, I therefore intended to look at design practices from the lens of manipulation, and not the lens of persuasive design. This study helped us in two ways. First, to establish the importance of disentangling what vulnerability online means. Second, to understand the tensions designers face when designing and to identify pain points in their understanding of what users' agency or harm constitutes, which ultimately contributes to users' vulnerability. These tensions have been useful in drawing conclusions and recommendations.

RQ2. How can HCI contribute to the conceptualisation of vulnerability to manipulative designs? This question contributes to conceptualise the problem of the experience with manipulative designs as a cause of vulnerability. Building on theories of perception and interaction, I explain how social and contextual differences in users contribute to how users interact with manipulative designs, giving rise to drivers of vulnerability. This theoretical conceptualisation anchors the dissertation and sets the frame for empirically investigating the drivers of vulnerability.

RQ3. What are the contextual aspects that drive vulnerability to manipulative interfaces? By addressing RQ3 we provide empirical insights into the experience of manipulation and a set of preliminary drivers of vulnerability. These expand the theoretical conceptualisation of Malgieri's conception of vulnerability drivers into empirical drivers stemming from users' ecologies and contributes to the experience of manipulation. Understanding what contextual drivers render users vulnerable to manipulative designs, allows us to identify design

challenges and to ideate on intervention spaces to design counter interventions.

RQ4. What are intervention spaces to design mitigation strategies for manipulative designs that account for vulnerability? To address this question, I conduct a rapid review and mapping of current intervention spaces suggested in the literature on design and HCI. Based on a re-analysis of qualitative data collected in Chapters 7 and 8, I identify drivers of vulnerability and use them as a basis for ideating new intervention spaces and strategies. I also discuss the design challenges brought by the drivers of vulnerability to support the design of countermeasures to manipulative designs.

2.3.1 Contributions of this Dissertation

This dissertation provides several contributions to the understanding of vulnerability towards manipulative designs for the HCI and design research communities. Given the interdisciplinary nature of the topic, this work also offers, to a lesser extent, opportunities for other disciplines to learn from the insights presented, including the legal community and policy-makers.

The first major contribution of this work is the empirical investigation of vulnerability towards manipulative designs. The empirical studies conducted contribute new knowledge about how vulnerability is manifested in different situations, contexts, and populations. Chapters 7, 8, and 9 unveiled specific drivers and contexts where vulnerability towards manipulative designs manifests. Although these drivers have been extracted from specific populations traditionally categorised as “vulnerable”, the importance of this contribution resides in our hypothesis that these drivers are not tied to specific populations but to anyone who likely shares the same context and experiences. For that reason, I categorise them as *experiential drivers* of vulnerability: social, interaction and agency drivers. As an example later discussed, the lack of feeling of agency or self-efficacy is a driver identified in older adults, which can be shared by other users for reasons likely unrelated to age. Thus, this dissertation contributes to moving away from labels that might stigmatise populations and look beyond these, which becomes crucial to defining design counter-interventions.

As a design-led major contribution, this dissertation identifies challenges and opportunity spaces for the design community to design counter-interventions for manipulative designs to prevent users from experiencing situations of vulnerability. Building on the drivers of vulnerability suggested in Chapter 10, I elicit design challenges and interventions that may help to mitigate vulnerability drivers and, therefore, contribute to ‘disable’ the effects of manipulative designs as triggers for vulnerability by accounting for vulnerability as a design material. Chapters 7, 8, and 9 provide contextual and personal drivers of vulnerability and open new design spaces for counter interventions — like social or recovery spaces for example. Ultimately, by using vulnerability as a lens to understand manipulative designs, I support researchers in identifying changes in the experience that call for different types of interventions.

As a methodological contribution, this dissertation offers some methodological reflections for the study of manipulative designs. The study of online manipulation as “something users cannot perceive”, as explained in Chapter 4, raises methodological challenges and underly-

ing reflections. Relying on different methodologies to overcome some of these barriers, we engage with this topic in Chapter 11 by offering methodological reflections for the research community, calling for the next wave of research on manipulative design. Hence, this dissertation brings together knowledge that other areas of HCI research brought. This aims to face similar challenges of situated experiences, helping to generate new threads of research in the area of manipulative designs research, and helping to restate the focus of such area. By studying what vulnerability means in the context of manipulative designs, this thesis has identified some opportunities for expanding the work on manipulative designs in a way that can protect users. It is not the merit of this thesis to bring approaches and methods that already have a long tradition in HCI and social sciences, but to see their value for the research on manipulative designs and suggest a roadmap on how to apply them in such a context.

2.4 Thesis Overview

Part I of the dissertation sets the motivations for this work and research approach, already presented in **Chapters 1** and **2**. Representing the state-of-the-art of the field, **Chapter 3** provides a critical review of the literature on manipulative designs. I explain the experiential aspects known in the literature and the main design interventions deployed in prior work.

Chapter 4 describes the methodological approach according to the research objectives and research questions. It offers an overview of the data collection methods used in the different studies, and the rationale for such choices. It also details the *raison d'être* of the collaboration with different organisations to recruit participants, as well as the approach to care embedded in the methodological design. Finally, it provides a reflection on my positionality and perspectives in relation to this work.

Part II of the dissertation focuses on situating the problem. It explains how manipulation can be engineered and how the relationship with one's environment can contribute to vulnerability to manipulative designs. **Chapter 5**, therefore, explains the tensions that UX practitioners encounter to avoid designing manipulative designs — addressing RQ1. Some of these tensions result in dangerous practices that can trigger vulnerability. **Chapter 6** addresses RQ2 by providing a theoretical analysis of what vulnerability to manipulative designs means in legal texts and how these concepts can be better nuanced through HCI theory. Relying on HCI theories and tools, this chapter demonstrate the value of investigating the situated experience of users with manipulative designs to understand their vulnerability.

Part III of the dissertation focuses on the exploration of experiences to identify drivers of vulnerability, contributing to RQ3. By investigating different groups of users and their relationships with manipulative designs, this part unveils some experiential drivers of vulnerability, which we hypothesise can be extended to users who share similar ecologies and experiences. **Chapter 7** explains the experience towards manipulative designs of teenagers at social exclusion risk. It provides new design spaces and challenges: designing for the social aspect of manipulative designs. **Chapter 8** showcases the experiences with manipulative designs in older adults. Using the magic machines method [11], it describes older adults' imaginaries

of manipulation and underlying imagined resistance strategies. It also unveiled challenges to design counter interventions: designing for lack of self-efficacy and risk aversion. **Chapter 9** provides insights into how users with lower digital skills understand and resist manipulative designs, helping to understand the problem of normalisation of these manipulative practices as a precursor of resistance practices.

Building on all gathered empirical insights, **Part IV** discusses the implications of looking at the users' experiences with manipulative designs through the lens of vulnerability. Thus, **Chapter 10** explores potential intervention spaces to address vulnerability to manipulative designs. By mapping the user experience in the presence of manipulative designs, this chapter contributes to rethinking existing intervention spaces and including new ones that leverage vulnerability to protect users with their own resources. Lastly, **Chapter 11** discusses how a focus on experience supports the understanding of vulnerability drivers to manipulative designs and their implications for the design community that aims to protect users from manipulative designs. The dissertation ends with a call to action for different stakeholders and the research community in manipulative design to broaden the understanding, methodologies and lived experiences of users within the interaction with manipulative practices.

2.4.1 Publications Included in this Dissertation

The following peer-reviewed publications are included in this cumulative doctoral thesis:

1. **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. Proceedings of the 2023 ACM Designing Interactive Systems Conference, 2408–2422. <https://doi.org/10.1145/3563657.3596013>
2. **Lorena Sánchez Chamorro**, Carine Lallemand, and Colin M. Gray. 2024. “My Mother Told Me These Things are Always Fake”—Understanding Teenagers' Experiences with Manipulative Designs. Proceedings of the 2024 ACM Designing Interactive Systems Conference, 1469–1482. <https://doi.org/10.1145/3643834.3660704>
3. **Lorena Sánchez Chamorro**, Romain Toebosch, and Carine Lallemand. 2024. Manipulative Design and Older Adults: Co-Creating Magic Machines to Understand Experiences of Online Manipulation. Proceedings of the 2024 ACM Designing Interactive Systems Conference, 668–684. <https://doi.org/10.1145/3643834.3661513>
4. Colin M. Gray, **Lorena Sánchez Chamorro**, Ike Obi, and Ja-Nae Duane. 2023. Mapping the Landscape of Dark Patterns Scholarship: A Systematic Literature Review. In Designing Interactive Systems Conference (DIS Companion '23), July 10–14, 2023, Pittsburgh, PA, USA. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3563703.3596635>
5. Arianna Rossi, Rachele Carli, Marietjie W. Botes, Angelica Fernandez, Anastasia Sergeeva, **Lorena Sánchez Chamorro**. 2024. Who is vulnerable to deceptive design patterns? A

transdisciplinary perspective on the multi-dimensional nature of digital vulnerability. *Computer Law & Security Review*, 55. <https://doi.org/10.1016/j.clsr.2024.106031>.

6. **Lorena Sánchez Chamorro** and Carine Lallemand. 2024. Towards a Second Wave of Manipulative Design Research: Methodological Challenges of Studying the Effects of Manipulative Designs on Users. CHI 2024 Workshop on Mobilizing Research and Regulatory Action on Dark Patterns and Deceptive Design Practices. <https://ceur-ws.org/Vol-3720/paper4.pdf>
7. **Lorena Sánchez Chamorro**. Perceptions of Manipulation and Resistance Among Low Digitally Skilled Users. 2024. *Submitted for journal publication*.
8. **Lorena Sánchez Chamorro**, Romain Toebosch, and Carine Lallemand. 2024. Mapping the Experience of Online Manipulation: Rethinking Intervention Spaces and Resistance Strategies Against Manipulative Design. 2024. *Submitted for journal publication*.

The following publications have not been directly included in the manuscript but contain insights that my thesis builds on and that have shaped my research process altogether. These publications are cited in the manuscript:

1. **Lorena Sánchez Chamorro**, Kerstin Bongard-Blanchy, and Vincent Koenig. 2022. Justice in Interaction Design: Preventing Manipulation in Interfaces. Position Paper presented at the Workshop "Dreaming Disability Justice in HCI". CHI Conference on Human Factors in Computing Systems (CHI'22), April 30–May 05, 2022, New Orleans (Virtual Workshop). ACM, New York, NY, USA, 3 pages.
2. **Lorena Sánchez Chamorro**. 2023. Disentangling Online Manipulation Strategies from the Perspective of Digital Inequalities. In Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23), April 23–28, 2023, Hamburg, Germany. ACM, New York, NY, USA 4 Pages. <https://doi.org/10.1145/3544549.3577060>
3. Cristiana Santos, Arianna Rossi, **Lorena Sánchez Chamorro**, Kerstin Bongard Blanchy, and Ruba Abu-Salma. 2021. Cookie Banners, What's the Purpose? Analyzing Cookie Banner Text Through a Legal Lens. In Proceedings of the 20th Workshop on Privacy in the Electronic Society (WPES '21), November 15, 2021, Virtual Event, Republic of Korea. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3463676.3485611>

Related Work: Mapping the Landscape of Manipulative Designs Scholarship Work

The first part of this chapter is based on the following peer-reviewed publication:

Colin M. Gray, Lorena Sanchez Chamorro, Ike Obi, and Ja-Nae Duane. 2023. Mapping the Landscape of Dark Patterns Scholarship: A Systematic Literature Review. In Companion Publication of the 2023 ACM Designing Interactive Systems Conference (DIS '23 Companion). Association for Computing Machinery, New York, NY, USA, 188–193. <https://doi.org/10.1145/3563703.3596635>

Abstract. Dark patterns are increasingly ubiquitous in digital services and regulation, describing instances where designers use deceptive, manipulative, or coercive tactics to encourage end users to make decisions that are not in their best interest. Research regarding dark patterns has also increased significantly over the past several years. In this systematic review, we evaluate literature (n=79) from 2014 to 2022 that has empirically described dark patterns in order to identify the presence, impact, or user experience of these patterns as they appear in digital systems. Based on our analysis, we identify key areas of current interest in evaluating dark patterns' context, presence, and impact; describe common disciplinary perspectives and framing concepts; characterize dominant methodologies; and outline opportunities for further methodological support and scholarship to empower scholars, designers, and regulators.

3.1 Introduction

Tactics of technology manipulation — often described through the concept of “dark patterns” [169] — are increasingly ubiquitous in digital services, and regulators are beginning to act in banning the most aggressive practices under consumer protection and data privacy law [56, 284]. Almost since the beginning of this term being coined in 2010, HCI scholars and practitioners have been central to this discourse, an area that is only rising in prominence and volume of research publications. As Lukoff [282] reported in a 2021 CHI workshop on “dark patterns,” scholarship on the topic is quickly rising within the HCI community and beyond, from a small trickle in the mid-2010s to more than two dozen publications per year starting in 2021. Obi et al. [339] similarly reported a rise in conversations about dark patterns on Twitter over this time period, demonstrating that not only is the issue of dark patterns of increasing interest and concern, but this discourse also involves a broader range of stakeholders over time—beginning with designers and technologists and now including social scientists, computer scientists, journalists, regulators, and law scholars.

While scholarship on the topic of dark patterns is quickly increasing, it is unclear which types of deceptive design practices need to be better understood, which types of patterns produce the most harm, and what kinds of studies regulators and legal professionals need to effectively identify, characterize, and sanction the use of dark patterns in technology systems. In this work-in-progress paper, we examine prior scholarship on dark patterns—including the breadth of methods and contexts employed and common framings and disciplinary motivations for studies—allowing us to identify opportunities for new research that both extends the state of the art and produces action in the form of practitioner guidance and regulatory sanctions. Additionally, our description of existing methods, contexts, framings, and contributions used in studies provides a pathway for scholars new to the space to explore avenues for contribution while also allowing existing researchers to better understand this emerging area of study. Our contributions in this PWIP work are two-fold. First, **we describe a work-in-progress analysis of common framings of dark patterns scholarship**, including dominant methodologies, contexts, and disciplines. Second, **we offer provocations for future scholarship opportunities, gaps, and areas of tension**.

3.2 Method

The methodology of this systematic review was guided by recommendations from the PRISMA report [313]. We used the following procedure to identify and screen literature to include in our analysis (summarized in Figure 3.1). All stages were carried out by a research team comprising six members that had previously engaged in dark patterns scholarship including investigators located in the United States, Luxembourg, and India.

3.2.1 Literature Collection and Screening

We conducted a search on Google Scholar and the ACM Digital Library using the search string “dark patterns” (with quotation marks) on September 13, 2022. This search returned 6,810 results in Google Scholar, and 183 results in the ACM DL (sorted by “relevance” in both databases) with no filters used to limit the search results. We downloaded the first 249 results from Google Scholar¹ and all 183 results were downloaded from the ACM DL, all as BibTeX. We then loaded all BibTeX entries into Rayyan.ai², a tool for collaboratively conducting systematic literature reviews. Based on our review of the resulting 432 titles and abstracts, 64 duplicate records were identified and removed. In addition, 25 records that were not written in the English language were removed. Finally, we attempted to download the full text of the remaining 343 reports and were able to retrieve all but two reports.

We then screened the remaining reports for eligibility. The following four inclusion criteria were: 1) the record had to be written in English; 2) the record had to mention “dark patterns” explicitly in the full text; 3) the record had to be published in a journal, conference proceedings, government technical report, or similar archival venue; and 4) the record had to include at least one empirical component. We excluded reports based on the following factors: reports that used an ineligible publication type (e.g. popular press article, only an abstract, workshop paper, student thesis, preprint; n=89); reports unrelated to computing or dark patterns (e.g., medical or hard science publications; n=17); reports that did not use dark patterns as a primary analytic or conceptual framing (e.g., only referencing the term in an introduction or conclusion; n=131); and reports which were not empirical in their framing (e.g., no new data collected or analyzed; only focused on argumentation; n=25). These criteria resulted in the exclusion of 262 reports from this review. This procedure resulted in 79 reports that were eligible for the analysis phase of our review.

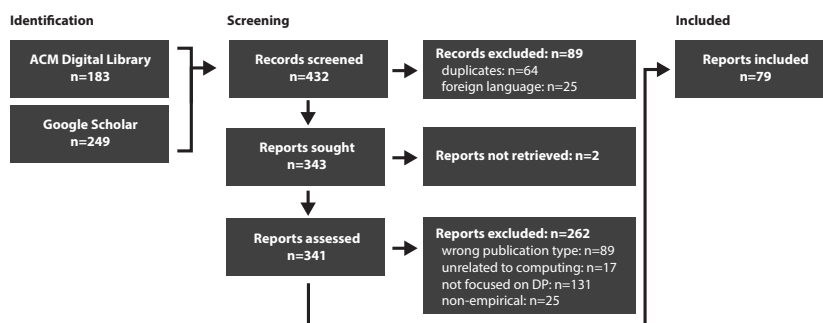


Figure 3.1: Systematic review flow diagram describing our identification and screening process.

¹We may have missed additional relevant literature beyond the first 249 results but after the first 150 results, few results showed evidence of citations and the vast majority of records appeared to not be relevant to the search. Future work could address this broader swath of results and indicate if we missed any important literature.

²<https://www.rayyan.ai>

3.2.2 Literature Analysis

To analyze the included literature, we used inductive qualitative content analysis [208] to answer the following questions: 1) What is the dominant context being addressed?; 2) What method(s) are used to empirically investigate instances of dark patterns?; 3) What framing concept(s) are used to motivate the study of dark patterns; 4) What is the publication type?; and 5) What appears to be the primary field for the intellectual contribution? To achieve this goal, four researchers on the team read the abstracts and relevant portions of the reports to ensure inclusion and code the selected records. The coding dimensions varied depended on the record and its contents. For example, a paper may contain more than one study on dark patterns and include varied evaluation methods and contexts. We began with a preliminary round of codes generated by the principal investigator based on an initial review of more than half of the dataset with sub-codes for each area of evaluation. This analysis was then supplemented by a review of five randomly distributed reports in the dataset by four other researchers to confirm the relevance and breadth of the preliminary codes, resulting in the addition of multiple sub-codes. Initial analysis work informed our creation of a codebook (Table 3.1), which we then used to code the entire dataset of 79 reports. One researcher coded all reports and then confirmed with at least one additional researcher. The principal investigator also provided consistency checks of codes across the entire dataset, engaging in conversation with the other researchers to reach full agreement on all application of codes. All categories of codes were applied non-exclusively except for the “framing” codes which we applied exclusively. Our final code application and codebook are included as supplemental material to extend our work.

3.3 Findings

We describe the outcomes of our analysis in relation to the context(s) addressed in the reports, the use of methods to support scientific inquiry, the framing and related contribution of the studies, publication type, and relevant disciplinary perspective of the publication venue. We report on the sub-codes we identified and focus on trends and gaps that we observed, supporting provocations for future work in Section 3.4.

3.3.1 Context(s) Being Studied

There was a broad diversity of contexts of study described in numerous ways, which we characterized as *specific functionality or domains* and *genres of sites or services*. The current landscape of research on dark patterns focuses disproportionately on genres where interaction occurs (e.g., social media, games, e-commerce; n=67) compared to precise indications of interface functionality, domains of use, or framings of interaction that include the use of dark patterns (e.g., consent banners, subscriptions, social robots; n=22).

Table 3.1: Codebook used by the research team to conduct confirmatory coding on the final dataset (n=79).

Context	
<i>Genres of sites/services</i>	
Advertising	any aspect of advertising platforms or advertisements
Digital Services	any website or service not described by any of the other types
E-Commerce	websites or apps used to purchase goods or services
Games	mobile or desktop games or game platforms
Mobile Apps	any apps on a smartphone or tablet
Social Media	any social media service (e.g., Facebook, Instagram, Tiktok, YouTube)
<i>Specific functionality or domain</i>	
Consent Banners	consent banner components as required by the GDPR or other regulatory statute
Physical Computing	elements of interaction in physical space and/or brain-body interfaces
Social Robots	elements of interaction with a social robot
Subscriptions	the process to sign up for or cancel a service
Voice User Interfaces	elements of interaction with a VUI
Methods	
Experimental	user behavior is characterized based on a comparison of control and experimental groups
Observation	user behavior is collected and tracked in real time, virtually or in person
Interview/Focus Group	a structured or semi-structured interview or focus group is conducted
Survey	user behavior or responses are collected through an online survey
Content Analysis	interface elements are evaluated qualitatively or/and quantitatively
Web Measurement	web sources are scraped and code evaluated, with some combination of manual and automatic inspection of the DOM or other code elements
Case Study	evaluation of a specific existing or speculative artifact that is used to support broader argumentation

Continued on next page

Table 3.1 – continued from previous page

Literature Review	systematic or otherwise disciplined review and synthesis of previous literature
Design	creation of new design alternatives to support downstream studies or as a means of speculative engagement
Diary Study	collection of data regarding technology usage or experiences over time
Framing	
Evaluative	leveraging existing taxonomies to identify whether something is an example of a dark pattern
Descriptive	using examples to illustrate power, impact, or attributes
Detection-Focused	creating and deploying an automated detection technique
Taxonomy-Building	defining new types of dark patterns or consolidating existing patterns
Problematizing	identifying limitations of taxonomies, identifying gaps in current literature
Experimental/Causal	identifying generalizable causal mechanisms

- *Genres*: Our dataset included studies focused on many different genres of sites and services. Genres included *games* (n=7), *advertising* (n=2), *travel services* (n=2), *social media* (n=15), *e-commerce* (n=6), and *mobile apps* (n=7). Many studies also included a mixture of many different kinds of web services or genres which we coded as *digital services* (n=27).
- *Specific Functionality*: Studies were also framed based on the presence or impact of dark patterns in specific moments of interaction (e.g., consent banners), interactive flows (e.g., subscriptions), and interaction domains (e.g., social robots, VUIs, physical computing). By far, the most common functionality-driven approach to studying dark patterns was *consent banners* (n=16). *Subscriptions* were far less commonly studied with only one contribution. Emerging domains of interest in relation to dark patterns were also infrequently studied, yet likely to be important in future dark patterns research, including studies involving *voice user interfaces* (VUIs; n=1), *social robots* (n=2), and physical computing interactions (n=2) through lenses such as proxemics and brain-computer interaction.

3.3.2 Method(s) Used to Study Dark Patterns

The most common method used to study dark patterns was *content analysis* (n=36), through which scholars evaluated interface elements and identified whether they contained dark patterns or not. In some studies, content analysis was central to the study, but in other reports,

content analysis was used to identify the presence of dark patterns that is then used to frame an experiment. The second most common method was *experimental* (n=14), in which the study included control and treatment groups to study the effects of particular design elements on user behavior. This method was particularly common in the context of consent banners, with half of the reports focusing on consent banner elements' impact on user behavior. The third most common method was *surveys* (n=10), used to gather users' behaviour, attitudes or opinions. *Literature review* of previous studies (n=8), the use of *case study* methodology to deeply interrogate the presence and/or impact of dark patterns (n=8), *design*-focused methods that involved the creation of new artifacts (n=7), *observations* (n=6), *interviews* (n=6), and *web measurement* (n=5) represent the remaining methods used in the reports we analyzed. The majority of studies attend to the existence of dark patterns through evaluation of the constituent elements in user interfaces, with less attention to the effects of dark patterns on users. Although these approaches might help policymakers and agencies in their consumer protection investigations, it may not contribute to a better understanding of the impact of dark patterns on users.

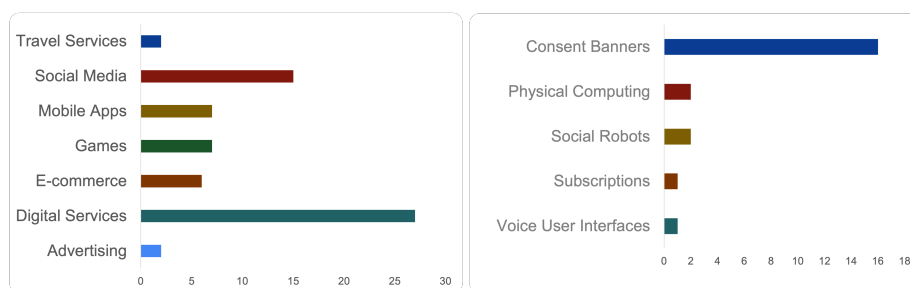


Figure 3.2: Distribution of papers by genre (top) and specific functionality (bottom)

3.3.3 Framing of the Contribution

Reports included differing framings representing different methodological goals and limitations, which reflected the primary intent of the research contribution. A *descriptive* framing, using examples to illustrate the power, impact, or attributes of dark patterns, was the most prominent in our dataset (n=19). Notably, reports with this framing were often addressing issues in new contexts or through new approaches, and many of these reports utilized case studies or content analysis to support their contribution. A *problematizing* framing, identifying limitations of the taxonomies and finding gaps in the literature was the second most common approach (n=16). These reports also tended to use qualitative approaches, such as interviews, case study methodology, or content analysis to extend the literature base and identify areas for further study at scale. Reports with an *experimental/causal* framing sought to identify generalizable causal mechanisms relating to the presence of dark patterns, representing the third most common framing (n=15). Unsurprisingly, the vast majority of these studies used an experimental methodology to support their claims, with some reports also using observational

or content analysis approaches. Reports with an *evaluative* framing (n=14) leveraged existing taxonomies of dark patterns to identify the presence of dark patterns in a new context with notable examples in social media, mobile apps, and consent banners. Reports with a *taxonomy-building* framing (n=12) focused on defining new types of dark patterns or consolidating existing patterns, with about half tending to explore a range of new contexts or lenses, such as privacy, games, social media, e-commerce, and physical computing. Reports with a *detection-focused* framing (n=3) were the least common, describing the creation and deployment of an automated detection technique. Altogether, the diversity of framing contributions within the dark patterns literature revealed particular dimensions of dark patterns scholarship that could be used to inspire future research studies within and across these framing categories.

3.3.4 Contribution Type and Field

Conference publications were the most common, with 59 articles published at various conference venues, while journal publications (n=19) and a workshop paper (n=1) were less common. The articles published at these venues came from diverse disciplinary perspectives. Although we coded only at the report level and did not investigate individual authors, HCI venues were the most dominant at 40 articles. Other contributing disciplinary venues included Game Studies (n=5), Privacy and Security (n=4), Computing (n=20), Communication (n=1), Human Factors (n=1), Design (n=1), Marketing (n=1), Law (n=1), Tourism (n=1), Business (n=1), Rhetoric (n=1), and Others (n=5). These disciplinary perspectives implicitly surface the range of fields and knowledge bases that have been potentially impacted and interested in studying the manifestations of dark patterns. These articles were submitted to diverse—and generally high-prestige venues—including HCI venues with high impact potential (e.g., CHI (n=17), DIS (n=4), CSCW (n=6)) and other computing venues that show the growing reach of dark patterns scholarship across international and non-US/EU populations (e.g., contributions in Brazil and India) and in relation to other computing perspectives (e.g., privacy and security; computing education; information systems; health informatics). However, because few studies reported the populations that they studied, it is difficult to assess specific trends—although it is clear that future research should embrace greater population diversity to address regional and culturally-specific manifestations of dark patterns.

3.4 Provocations to Grow and Support a Future Landscape of Dark Patterns Scholarship

Building on our analysis of the current empirical landscape of dark patterns scholarship, we are able to show not only the breadth of study designs, contexts, and framings, but also emerging consensus on best practices to continue to grow scholarship in this important area. In this section, we provide a set of provocations for future work to extend guidance on dark patterns. These provocations outline some emergent challenges and areas of scholarly consensus, and

indicate some qualities of future scholarship that could increase the translation of efforts among scholars, design and technology practitioners, regulators, and legal professionals in the battle against dark patterns.

Dark Patterns Scholars Should Build Alignment and Community Norms. The diversity of the scholarship we analyzed—across differing methodological traditions, contexts, and fields of study—is a strength, revealing the harms of dark patterns in many of the systems that define our technologically-mediated existence. However, this diversity also comes with potential risks if we do not find areas of alignment and empirical norms. We prefer to think of these risks as *opportunities* rather than explicit gaps, since considerable and converging scholarship has already laid the groundwork for the empirical challenges that come next. First, as observed by Mathur et al. [297] and others, there is a lack of shared vocabulary which could be constructively addressed through the creation of a shared ontology to align scholarship across type and discipline. First steps towards this goal have been recently published by Gray, Santos, and Bielova [172], but substantial work remains. Second, we observed methodological pluralism across the landscape of scholarship which is a sign of strength, but within specific context and methodology categories, validity or reliability threats were present in some studies. For instance, there was variable quality in arguments derived from literature review-focused studies, and experimental studies also included differing levels of rigor depending on how they operationalized dark patterns (and at what level of granularity). To further support ecological validity, future work should also identify the kinds of demographic characteristics that should be collected for replication and shared knowledge building, alongside an evaluation of the ecological validity of various methods to address specific detection challenges. Third, we found distinct silos of different study framings but few connections across these silos. For instance, are there better ways for scholars to connect evaluation, problematization, and detection-oriented work? If we can solve this challenge, it could increase the translational effectiveness of the work, including new opportunities for transdisciplinary engagement and social impact via legal and regulatory channels that can leverage knowledge built through empirical studies.

Dark Patterns Scholars Should Outline Methodological Strengths and Weaknesses. We observed clear “types” of papers that also point to dominant methodological traditions. These types include: *exploratory work* to identify what dark patterns look like in specific places, naming their presence; *experimental work* to characterize the differential impacts of dark patterns building on existing typologies and exploratory work; and *detection-oriented work* that describes the presence of characterizable dark patterns at scale. Within each “type,” there are common combinations of methods used, but there are also areas where more specific methodological guidance will be important to support the next wave of scholarship. For instance, researchers must be able to identify and characterize the presence of a dark pattern, which is almost always achieved through content analysis or other similar expert inspection technique. Specific standards of reporting will be important to allow this inspection and characterization to be falsifiable and encourage a shared language regarding the presence of dark patterns in specific, situated instances. It is also important for studies across these three

types to engage in conversation with each other, including key methodological limitations. For instance, detection work has only been productive for very specific patterns and types of functionality (e.g., consent banners, e-commerce) and future work may need to include more explicit human-in-the-loop techniques to overcome the lack of detectability of many pattern types. Similarly, experimental work should leverage and acknowledge the ecological complexity of patterns in-the-wild (often addressed in exploratory work) and not seek to diminish or reduce that complexity in ways that undermine ecological validity, while also seeking to describe the power of dark patterns on users.

Dark Patterns Scholarship Should Offer Translational Opportunities. Although our focus was on empirical studies, we acknowledge the need to involve many stakeholders that benefit from this knowledge. As dark patterns are increasingly under legal scrutiny, how do we effectively translate and activate dark patterns knowledge across academic research, practitioner, regulator, and legal professional communities? One area of engagement that we left unexplored in this analysis could include emergent conversations in the legal community regarding dark patterns; due to differences in knowledge generation approaches, this argumentation-focused work (e.g., [267]) was excluded from our analysis in this paper but could provide cues towards transdisciplinary cooperation at the intersection of HCI and law (e.g., [181]). Other practices that could increase translation across communities could include leveraging shared vocabulary to indicate which specific type(s) of dark patterns are being studied, in what contexts or with what type(s) of users, and with what assumptions of the underlying definition of dark patterns. Our work could also identify common elements of reports based on the framing goal of the contribution to increase opportunities for meta-analyses of key areas of interest.

3.5 Towards Understanding Users Experience with Manipulative Designs

This section provides a critical review of the existing literature through the lens of the experiences with manipulative designs. I provide an overview of the main aspects related to the experience with manipulative designs, as well as some of the factors that have been studied in the HCI and design literature. Lastly, I also provide the main interventions that have been discussed within the community.

A methodological note on this synthesis. This section builds on the findings (and databases) from the aforementioned systematic literature review from Gray et al. [171], as well as the more recent scoping review from Seaborn and Chang [393], which already extended the former. Considering the time frame in which the data from both studies was collected, I have expanded the dataset with papers dated until 1st June 2024. While this is not a systematic review, the papers included can be found in the Supplementary Material ³ for transparency purposes. It is important to mention that, in line with the aforementioned reviews, the database

³bit.ly/supplementary-material-rapid-review

selected to extract new material is ACM Digital Library. This selection is made, not only because of coherence with the authors, but also because this thesis intends to review applied-oriented work following the nature of the HCI and design community. Therefore, some work that is done outside the HCI domain, will remain outside the scope of this synthesis.

I here report the critical review by main themes. First, I summarise what experiential aspects have been analysed by the literature through the variety of methods presented in Section 3.3. Second, I briefly describe the type of sampling that has been studied. And last, I describe the different interventions suggested in the literature. This information helps to understand the gaps in the literature and which existing knowledge can serve as building blocks towards understanding manipulative designs in a broader experiential landscape.

3.5.1 Experiential Aspects

Recognition and Awareness of Manipulative Designs

Several studies aimed to understand awareness of manipulative designs among users. Researchers usually do it through identification tasks; hence, these studies focus on the recognition of manipulative designs. For users, identifying what the research community is categorising as manipulative designs is sometimes hard, what Di Geronimo [111] coined as “blindness”, a concept that was followed by other authors later on [308], and that explains the effect of blindness on users when exposed to manipulative designs [111, 308, 456, 230, 347, 411, 45].

Despite often not able to recognise them, users declare to be aware of the existence of manipulative practices online. When shown manipulative designs through a survey study, Bongard-Blanchy et al’s [45] participants were uncertain about whether the designs might influence them or not. Although there was general acceptability towards them, they were also aware that services might influence them in general terms, experiencing some harm. In their questionnaire study, Gray et al. [163] showed participants reported practices that make them feel manipulated: when they were asked about their privacy, when they feel threatened about their security, or they are asked to pay for products, participants feel manipulated. Thus, other studies have focused on more general user experiences when they encounter these designs [281, 78].

Prior work has tried to explain the reasons behind this ability (or lack of) to identify manipulative designs. With the assumption that awareness and identification of manipulative designs are related to the use and familiarity with technology, some studies have explored it. Bhoot et al. [286] found a positive association between the frequency of manipulative designs and perception of trustworthiness. However, Voigt et al. [453], by using the Affinity Towards Technological Interaction Scale [148], did not confirm the existence of an association between the affinity for technology and recognition of manipulative designs. Freeman et al. [150] have qualitatively explored the awareness of these designs by understanding users’ ethical concerns and perception of fairness in services with manipulative designs. Other external aspects have been questioned as potential reasons to identify, like familiarity with computer security and privacy concerns. Bermejo Fernandez et al. [34] found that participants who

self-reported a high level of familiarity with computer security and more self-reported privacy concerns, were more likely to change default designs on cookie consent banners than other participants.

Few studies have examined how socio-demographic variables affect the relationship with manipulative designs. Bongard-Blanchy et al. [45] and Luguri and Strahilevitz [280] showed how levels of education had an impact on recognising manipulative designs and resisting them. Similarly, Bongard-Blanchy [45] explained how older participants had more trouble finding manipulative designs, resonating with Avolicino's [15]⁴, who found older adults to be less acquainted with the service. Noteworthy, Zac et al. [472] did not find differences in effectiveness on manipulative designs between groups with different socio-demographic conditions — age, education and socio-economic status.

Perceptions of Resisting Manipulative Designs

The perception that users have about their capacity to resist resist manipulative designs — namely, self-efficacy — has been explored in some studies. While there is no unified established idea of “resistance” means, scholars consider “resistance” when users avoid falling for the manipulative design — regardless whether they actively seek this avoidance or not. Bongard-Blanchy et al. [45], observed a negative relationship between people's perceived influence to resist manipulative designs and their capacity to recognise them: people that recognise manipulative designs consider themselves more capable to resist them. These results resonate with Maier and Harr's [287]. In their study, when participants were shown these designs, they assessed them in relation to the potential impact on users, but also “accepted them” because they find them easy to ellude “by clicking the X”.

The role of users' perceived self-efficacy has gained importance in the literature. By investigating possible countermeasures, Schäfer et al. [389] discuss the dichotomy between the user's desire to remove excessive information from the interface — that might be misleading —, while they want to be in control of what is removed, demanding agency. In this line, Lukoff et al. [281] explored how to help users regain agency when facing deceptive attention capture patterns. Chaudhary et al. [78] and Lyngs et al. [285] made the same claims for giving users' agency when they face manipulative designs.

Some studies have explored the assumption that with more knowledge, users will be more likely to avoid the effects of manipulative designs during the interaction. Chen et al. [80] thus developed a tool that provides knowledge to users within the interaction. Through the evaluation of the tool with users, it proved useful to help users recognise manipulative designs. With the same rationale, Tjostheim et al. [434] created a board game aimed at increasing knowledge of these instances among teenagers, although it did not prove useful for such purpose when it was tested. Hence, it is still unclear from these research works, whether more knowledge about manipulative designs can help users. Additionally, Strycharz et al. [427] found that technical knowledge in consent notices do not impact the change of default settings to opt-

⁴We believe that Avolicino's results in relation to age must be taken into account cautiously since their oldest cohort is “over 35 years old”.

out for personalized ads. Testing this hypothesis on manipulative designs concretely, Klütsch et al. [237] did not find a relationship between providing knowledge to the user when the interaction takes place and user consent.

The Relationship Between Trust and the Experience with Manipulative Designs

Trust in the service is a recurrent underlying topic in the studies conducted on manipulative designs. Gray et al. [163] explained how trust in the service and the different elements that contribute to distrust in users are fundamental triggers of the experience of manipulation. Bhoot et al. [286] found a positive association between trust and “the physical appearance of the platform”. Tuncer et al. [439] explore “perceived benevolence” (i.e. the perception of someone’s good intent) as a sub-dimension of trust and showed a negative relationship between the presence of scarcity cues and the perceived benevolence of the services. Similarly, Mildner [306] explained participants need trust in the development of ethical conversational user interfaces; however, this development of trust can flip since, as Sanchez Chamorro et al. [379] describe, UX/UI designers want to build trust from users to influence their behaviours.

Technological Factors Impacting the Experience of Manipulative Designs

Although there are studies targeting specific technological modalities and devices — [306, 243], mobile applications [111, 180], web interfaces, or augmented/virtual reality [128, 456] —; a limited set has explored the device as a factor that affects the experience or behaviour in the presence of these designs. In their experiment, van Nimwegen and de Wit [447] see mobile users fall for manipulative designs more than desktop users. They discuss the idea of how mobile interfaces are designed in comparison to desktop ones, in which users can see more elements. Similarly, Bermejo Fernandez et al. [34] found that participants interacting via mobile phone had less time of interaction in comparison to the desktop ones, and that such interaction was uniquely to fall for the deceptive pattern and accept the cookie consent banner.

Emotions and psychological harm

Several studies have reported psychological harms that comes from the experience of online manipulation . Gray et al. [163], showed how the experience of manipulation was associated with negative experiences and emotions — distress, upset, guilt, fear, hostility, irritability or shame, among others — in line with the results of Avolicino et al. [15] and Hogan et al. [203]. Bhoot et al.[286] found that frustration was positively correlated with the identification of manipulative designs. Similar negative emotions of impulsivity were found in Chaudhary et al. [78] and discussed by Moser et al. [316]. Chaudhary et al. [78] also explains how participants show negative emotions and regret when they realise they have continued on the platform unintentionally. In the same line, the study of Tuncer et al. [439], on scarcity cues showed a relationship between negative emotions and the presence of these design elements.

3.5.2 Populations Studied

Within the literature, two main populations are represented: users and designers. While the studies that look at users try to understand how users behave in the presence of manipulative designs, the studies with designers try to understand how they create these interfaces and what their needs are if they are willing to avoid using manipulative practices in their designs, [166, 86, 83, 306]. When it comes to users, the majority of the studies do not focus on specific populations since their main intention is not to look at specific groups or to compare populations. Hence, there is a wide range of studies that rely on crowdsourcing platforms without specific sampling goals [31, 38, 45, 439, 34, 184, 281, 347, 390] or convenience sampling among authors' universities and networks [78, 281] — although sometimes this is not disclosed [15, 32, 46]. A few exceptions are studies “in the wild” that collaborate with real platforms to study end users interactions with the platforms' features and elements [336, 280]. There are also limited studies targeting teenagers' and kids' experiences with manipulative designs [140, 434, 369]. A subset of studies have also looked into digital communities with the aim of targeting specific populations like video game players [150]. Some other populations included in prior work are experts [306, 230] and university researchers as end users of manipulative design platforms [412].

Given a wide range of studies use Western-based crowdsourcing platforms, there is not a lot of variety in terms of geographical representation. This representation is more notable in evaluative studies that aim to assess a platform and its content in the search for manipulative designs. Thus, given the idiosyncrasy of some contextual elements belonging to different societies, countries, and contexts, some studies have looked at manipulative designs in specific countries [201, 401, 254, 119].

3.5.3 Interventions to Fight Manipulative Designs

Different types of interventions to counteract manipulative designs have been proposed using various approaches. The highly context-dependent and temporal aspect of the effects of manipulative designs makes it easy to understand the challenges of designing interventions that effectively work. Thus, to rethink design interventions, it is first necessary to understand the experiences of these manipulative designs as a dynamic and long-term process. Here, I refer to interventions as any action conducive to protecting users from manipulative designs. Hence, they depend on which level they act on: within the interaction between the user and the interface — e.g., UI interventions —, or around such interaction — e.g., regulatory or educational interventions.

Regulation and co-regulation

In this section I gather those studies that call for more binding regulations from governmental actors, as well as other mechanisms that impose a regulatory framework that companies would have to adapt but adapted to their own contexts, like codes of conduct and standards,

which is called co-regulation [365, 59, 67]. A huge number of studies call for more regulation and policymaking interventions [390, 45, 31, 347]. While most of them do it in a generic way since their primary focus is not discussing regulations, others suggest more specific approaches like enforcement on popular sites, banning known manipulative practices [458], and for specific modalities [347], the truthfulness of legal disclosures, requiring feature parity, holistic investigation enforcement [180], specific regulation for consent platform managers [336], or the use of internal governance mechanisms that empower users to avoid manipulative practices [180, 379], like standardisation [183], or templates [426].

Educational interventions

Some studies argue education as an intervention to create awareness of manipulative practices as it is meant to “favour users’ agency” [45] (p. 771) and suggest interventions meant to give users more information about how manipulative designs work by other mediums external to the interaction — i.e. educational programs or public awareness. These interventions usually refer to “education” in a broad sense [287, 45, 399, 230], by putting the emphasis on users need to be “educated” to increase their awareness on manipulative design, so they can, in turn, protect themselves from these practices. In a few studies, there are more specific suggestions about how this can be executed. Education about manipulation literacy [473], or creating tools that help awareness like plugins [241] and public websites with ‘metrics’ on the manipulative designs that products have [399]. Keleher et al. [230] more precisely suggest general campaigns on awareness that target daily technologies, since they believe can lead to collective action among users. On a more narrow perspective, Tjostheim et al. [434] have tested board games to increase the knowledge of manipulative designs among teenagers, and Klütsch et al. [237] tested cookie consent banners in which the information about manipulative design was provided; both types of interventions did not prove to be effective in such studies.

Measures targeting Designers and Developers

Prior work has discussed interventions that help designers and developers avoid the creation of manipulative designs during the design process. Scholars suggest the use of ethical frameworks like Value-Sensitive Design [180, 255], as well as rethinking and expanding existing ones with an emphasis on manipulative designs and their specific nature [169, 306]. Authors have also emphasized the importance of educating designers, in general [129, 169, 458], and concretely from the very early stages of their education [82, 83] as the basis for training future designers in more ethical practices. Other measures targeting practitioners involve educating providers about manipulative design [46], using design methods to test the presence of manipulative designs — e.g. A/B testing [314, 45] – within their design process. Specific technical tools for developers to prevent these practices are also suggested like APIs that helped them in their development [430]. Theoretical frameworks to help designers [70] and specific design guidelines and templates [261, 419, 426] are also suggested.

Technical interventions to automatise manipulative designs detection

Another set of interventions suggested are technical tools with the purpose of automatise the detection of manipulative designs [45]. This can ultimately have different subgoals: help regulators to detect and enforce [336, 236], help designers to develop [80, 279] or support users within their interactions [80]. Curley et al. [103] explained the complexity of certain manipulative instances made it complicated to automatise, while [241] suggested automatise tools with a can be community driven. Plugins that remove manipulative designs have also been explored by Monge Roffarello and De Russis [314] on Facebook. and Kolling et al., [241] and Twitter (now “X”).

UI and visual elements

Researchers have also tested potential UI interventions as differences in colour, positions or medium through the interaction take place — e.g. link vs button options should remain equal [31], rejection options need to be better signalled [32], using colour traffic likes schema to signal cookie options [34] Bielova et al. [38], in an experiment on cookie consent banners, included in their interventions a condition in which the risk of accepting was represented that proved to be effective in reducing cookie acceptance. Graßl et al. [162] also tested the effectiveness of “bright patterns” in the context of cookie notices, understood as the opposite of “manipulative designs”, namely UI strategies that steer users in favour of their own interests. They showed how using bright patterns helps to influence privacy choices — in this case, steering to privacy-friendly options.

By testing UI interventions in different digital services that use different settings — visual cluttering, information provided to users or positions—, Schäfer et al. [388, 389] discuss the dichotomy of users claiming more agency about the choices they make wanting more information about their decisions, but also feeling overwhelmed by the amount of information that some interventions provided.

Behaviour change strategies and friction

As Chang et al. [76] explain in their scoping review, there is a majority of studies in the community that use dual-theory as the main theoretical lens. Dual-theory Kahneman [223] refers to the existence of 2 different systems to process information—System 1 is automatic, while System 2 is reflexive. Hence, different studies propose design interventions that try to steer users away from the intentions of the manipulative instance by applying behaviour change knowledge trying to trigger the reflective system [45, 78, 129, 162, 173, 255, 441, 270, 237, 281]. Understanding the psychological aspects of manipulative designs [270], including social needs like the need to belong [47], to apply behavior change techniques is a common suggestion: using warnings, reminders, educative nudges within the interaction, [162, 173] and long-term boosters [237] is expected to trigger reflection when users interact with these designs. Lukoff et al. [281] and Lyngs et al. [285] suggest goal-setting techniques in the interface to mitigate attentional harms. By prompting goals that users have previously set, Lukoff et al.

[281] suggest reflection can be triggered and prevent users from spending excessive time on video platforms.

With this idea of increasing reflection, some studies conducted on general populations discussed the use of friction — i.e. as “are points of difficulty encountered during users’ interaction with a technology” [101] (p.2) — and micro boundaries — i.e. “intervention that provides a small obstacle prior to an interaction that prevents us rushing from one context to another” [101] (p.4) — as potential countermeasures [316, 45, 78, 281, 255]. Yet, very limited studies have tested it to date. Zac et al. [472] tested friction and proved useful on e-commerce payment.

Socio-technical interventions and others

Here I include interventions that do not necessarily relate to the interaction between users/de-signers with manipulative designs, but with the whole system in which manipulative designs exist. Chaudhary et al. [78] and Freeman et al. [150] advocate for interventions in the algorithmic system — i.e. varying recommendations, algorithmic fixing, and transparency. Power imbalances in the system are one of the main targets of social interventions [128, 150] which aim to look at users vulnerabilities [239], promote community and collective action [241], and foster users’ empowerment. Westin et al. [458] suggest empowering users against FOMO-centric design by accounting for the social aspect of privacy trade-offs and the need to belong [47]. In a more concrete approach, Keleher et al. [230] propose tip lines that can help users protect themselves, and Lu et al. [279] to include a citizen science approach in the study of manipulative designs, so users can be involved in the decisions about these designs.

Chapter 4

Methodology

This chapter is partially based on the following peer-reviewed publication:

Lorena Sánchez Chamorro and Carine Lallemand. 2024. Towards a Second Wave of Manipulative Design Research: Methodological Challenges of Studying the Effect of Manipulative Designs on Users. CHI 2024 Workshop on Mobilizing Research and Regulatory Action on Dark Patterns and Deceptive Design Practices. <https://ceur-ws.org/Vol-3720/paper4.pdf>

Abstract. The study of lived experiences in the context of a phenomena that is hard to recognise by users brings a wide range of methodological challenges. Building on such challenges, this chapter explains the epistemological decisions made to address the questions introduced in Chapter 2. I provide an overview of the selection of methods, as well as the approaches deployed to recruit participants. Lastly, I reflect on and describe my positionality as a researcher conducting this work.

4.1 Methodological Challenges in the Study of Manipulative Designs

While the body of knowledge on manipulative designs and their effects on users is growing, the highly context-dependent research problem entails specific methodological challenges — e.g., around validity [38, 31] and generalizability [442, 45, 336]. First, manipulative designs are embedded in a broader technological ecosystem that overlaps with other kinds of effects. For instance, where does the impact of the algorithm stop, and where does one of the manipulative design elements start? The existence of highly contextual effects — i.e. the trade-offs that users make depending on the time, place, or context of the interaction — multiplies the potential factors at play in the effect of manipulative designs, increasing the challenges for researchers.

Research is limited in this regard by the homogeneity of the contexts studied, the population sampled, and the methods used. A wide range of studies have been conducted in the context of cookie consent banners, but research on other contexts is lacking [171]. Similarly, the analysis of user interfaces by the researchers, without user involvement, has been the dominant research method used to date. In the literature review presented in Chapter 3 I identify some of these challenges and voiced a need for more methods and reflections. This dissertation identifies the following challenges inherent to the study of the effects of manipulative designs:

1. Contextual elements play a more significant role in manipulative designs, which are **intertwined with other platform affordances**, and their impact is mediated by individual, technical, physical, temporal, social, and task-related contextual factors.
2. These effects and potential harms of manipulative designs can take **several forms and are temporally situated**.
3. Online manipulation is a **phenomenon that (some) users are likely to be unaware of**, and yet impact and harm on users is the main consequence.
4. **Specific populations might be more impacted** by manipulative designs, and they are likely to be harder to reach when conducting research on technology.

In this section, I gather the main methodological limitations and challenges discussed by scholars in research papers on manipulative designs. Various methods from diverse perspectives — legal [181, 298], design [379, 324, 82, 86], or computer sciences [103, 296] — have been used to study manipulative designs, yet methods to understand the relationship between users and these designs are still limited [171]. A common method used is content analysis to evaluate the presence of manipulative designs on interfaces but does not necessarily involve end-users. As Gray et al. [171] explains, although the community is equipped with some exploratory methods and experimental protocols, there is a need for *“more specific methodological guidance to support the next wave of scholarship”* (p.192).

Without aiming to be exhaustive, I provide a brief overview of the challenges, limitations, and trade-offs reported by authors of empirical studies on manipulative designs involving users. I did not consider studies whose primary focus is on the design process, the existence of manipulative designs on a platform, or the characteristics of manipulative designs without considering users. Although the objectives of each study vary, it is relevant to compile the limitations reported by the authors to understand the challenges researchers are navigating in the study of the effect of manipulative designs on users.

Out of 79 studies on manipulative designs reviewed by Gray et al. [171], a majority (45%) evaluated manipulative elements in user interfaces through content analysis without the involvement of the users. Empirical studies focusing on the effects of manipulative designs on users were usually done through experiments (n=14) investigating the effects of particular design elements on user behavior [447, 38, 439, 433] or surveys (n=10) to gather users' attitudes, opinions, and reported behaviour [45, 286]. Other traditional methods such as observations (n=6) [15, 280, 286] and interviews (n=6) [1, 78] were used in prior work but to a lesser extent. I argue that the community needs more empirical research to better understand the experiences of online manipulation and the users' ecologies to inform design interventions. I review the methodological challenges reported in empirical studies on manipulative designs.

4.1.1 Translational Contexts and Validity Challenges

One of the biggest challenges in the study of manipulative designs is the search for ecological validity, to account for the situatedness of human action and contingency on contextual factors. Indeed, the specificity of every setting and context limits the transferability of insights into the effects of manipulative design. Bielova et al. [38] explained how their findings on cookie consent banners are hard to translate to other banners. The authors acknowledged their setting might be different from the usual settings in which users interact with consent banners, including the attentional limitations that users experience.

The fact that fake experimental settings do not embody the same consequences that the actual interaction on the user experience is a known limitation [257]. For this reason, Schaffner et al. [390] explained in their study limitations the impossibility of evaluating the impact of confirm shaming on their participants. Tuncer et al. [439] warn readers to take their results cautiously because the simulated e-commerce website used to test scarcity cues did not incorporate real financial incentives. Similar limitations are reported by van Nimwegen and de Wit [447]. To overcome this, Utz et al. [442] partnered with a company to study manipulative designs in a real context and Nouwens et al. [336] used a browser extension to test the manipulative designs in participants' everyday interactions. Moser et al. [316] envision industrial partnerships as a necessary step for future research on manipulative designs, given the complexity of the topic. I share some of their vision that underlies why ecological validity is key on this topic; while companies have the resources and real settings to test manipulative strategies on users as "attackers", researchers' resources as "defenders" are much more limited.

An additional problem arising from the dependency on context is the dissociation between the effects of manipulative designs and the rest of the platform's affordances. Bongard-

Blanchy et al. [45] explained that when asked about felt manipulation, participants mentioned other intertwined affordances instead. Therefore, several studies acknowledge their results to be valid only in the specific contexts and with the specific manipulative designs studied [38, 453, 447, 316].

4.1.2 Populations and Sampling

Another reported limitation relates to the population and sample strategies used. Many studies on manipulative designs use crowdsourcing platforms to collect data on users. While these systems have some merits in some contexts, their use in the context of online manipulation might decrease the study's validity and lead to inconclusive results. Berens et al. [31] explained that by trusting the study, crowdsourced participants would increase the acceptance of cookies. Many studies report sampling limitations. While trying to sample a broad range of participants' profiles, Bongard-Blanchy et al. [45] and van Nimwegen and de Wit [447] fell short in including older adults. Avolicino et al. [15] even presented their oldest cohort as "above 35 years old".

Several studies acknowledge limitations related to convenience samplings — via crowdsourcing platforms or recruitment within universities — as leading to the inclusion of an overall younger and more tech-savvy population, even when the researchers aimed for generalisation [31, 45, 316]. Noteworthy, with the few aforementioned exceptions that included some diversity in age ranges, the set of literature that includes the analysis of other variables as mediators of the effects of manipulative designs are scarce [280, 45, 447, 15]; and none of these included any population that embodies vulnerabilities, which is often labelled as "hard-to-reach" [398].

4.1.3 Observing the Harm versus Understanding the Harm

Many methodological challenges in our subfield seem to arise from the quest for generalizability, as a major part of the community adheres to quantitative research approaches. Yet it is worth reflecting on whether that path is an adequate starting point and which research objectives are meaningful. Scholars are divided between methods that aim to observe harm (understood as damage on users' lives [181]): a common quantitative approach (e.g., [31, 38]) opposes a qualitative view aims at understanding why and how those harms occur.

The literature review by Gray et al. [171] shows a wide use of qualitative methods from a designerly perspective. Examples of qualitative methods to understand users include Maier and Harr [287]'s focus group study, providing first insights into the perceptions of manipulative designs on users. Gray et al. [163] elicited the notion of temporality of felt manipulation via card sorting methods. Chaudhary et al. [78] diaries and interviews data unveiled the nuances embodied in the interaction with capture attention deceptive patterns. While these studies do not seek generalisability, they inform specific aspects of the relationship between users and manipulative designs in ways other research designs do not. Borrowing science and technology studies terminology, manipulative designs are relational [424]: they are only perceived

by the user when there is a negative impact. Gray et al. [169] already pointed out this phenomenon by defining manipulative designs as a mismatch between user expectations and the interface. Thus, many of the methodological challenges associated with this relationality.

Hence, there is a necessity to assess the methodological implications underlying the study of manipulative designs to advance the field and support practitioners and policymakers with a solid body of knowledge to inform countermeasures. Mathur et al. [297], in their review of challenges in the realm of manipulative designs, already advocated for the need to use design epistemologies to better understand users' interactions with manipulative designs. Similarly, the OECD [340] acknowledges the challenges of understanding vulnerability online as a situated phenomenon and calls for rethinking methods that can give a more nuanced understanding of the problem. Therefore, to investigate situations of vulnerability to manipulative designs, selecting methods that account for the lived experiences of participants is crucial.

4.2 Selection of Research Methods

Building on the aforementioned challenges, a main motivator in the selection of methods in this dissertation was to be able to understand the contexts of vulnerability to manipulative designs in a situated way while being able to understand users' perceptions of these manipulative instances. Thus, while a qualitative and experiential approach that helps to understand users' lived experiences is necessary, it is also essential to find a balance between extracting the perceptions of elements that are hard to perceive while staying true to the participant's experiences. With the objective of understanding lived experiences of participants, there is no epistemological commitment to establish causal relationships or produce generalisable insights [418].

The following chapters describe in detail different studies in which variations of qualitative interviews, and co-creation workshops have been applied. I hereby provide an overview of the methods selected to investigate the aforementioned research questions:

- **Interviews using critical incidents.** In Chapters 7 and 9, I used interviews following the critical incidents technique [373, 457, 259]. Participants were asked about "incidents" in their experiences with the internet. This technique allowed them to identify specific past experiences in which they felt manipulated and/or deceived.
- **Interviews with probes.** Complementing the critical incidents technique, I used probes representing manipulative designs during interviews. Here, probes are artefacts that represent something, in this case manipulative designs, in order to elicit participants' reflections and perceptions about manipulative designs. Therefore, this method is closer to the data collection use of probes [44], rather than cultural probes [158]. This combination of techniques aimed to find a balance between gathering their perceptions and comments about manipulative designs, while staying true to their situated experiences.
- **Co-creation workshops.** Chapters 5 and 8 use different formats of co-creation workshops, with the same purpose of eliciting experiences of the (sometimes) invisible phe-

nomenon of manipulation in different contexts. In the case of designers, in Chapter 5, workshops were meant as a tool to situate participants in a fictional task but inspired by realistic briefs that designers could encounter. In Chapter 8, the Magic Machines co-creation workshop was chosen to facilitate participants to talk about their situated experiences of technology-perpetrated manipulation, without the potential barrier to use technology [11, 7].

With the use of interviews with critical incidents technique, showing probes with manipulative designs, conducting task-oriented interviews, and the use of magic-machine workshops, I tried to overcome some of the aforementioned challenges. While this combination of methods also presents limitations (exposed in every chapter), it provides a nuanced understanding of the contextual experiences with manipulative designs in different ways.

4

4.3 Collaboration with Study Participants

The specific approach to vulnerability that underlies the objectives of this dissertation requires careful attention to care as a core value [249, 248]. Aware of the situations in which potential research participants could be, there was an extra duty of care. While this work cannot be categorised as participatory per se, I sought to follow the mindset of participatory design research: I aimed for participants to feel heard and get something in return for their participation [186, 63]. This duty of care was translated into different parts of the research process, from the engagement with participants to the debriefing information.

To engage with participants, for RQ3, I approached different organisations and NGOs that worked with a diversity of participants, from young adults at risk of social exclusion to older adults. Expecting an intersection of vulnerabilities to emerge among participants, I asked these organisations for support in the recruitment process to build rapport with my participants but also to ensure that participants were not affected by taking part in this research [108, 96, 448]. Creating trust was thus bidirectional, participants needed to trust the partner organisation, as they would make sure everything was adhering to the terms described in the Ethical Review Application to the University.

For the studies described in Chapters 7 and 9, I collaborated with “La Rueca Asociación Social y Cultural” (from now on “La Rueca”), an association based in Spain, with different offices in Madrid. This association works in socio-cultural interventions for young adults, teenagers, and kids at social exclusion risk. Among their activities, La Rueca provides “healthy” leisure time, support with homework, and formative activities to help young adults access the job market. For the study described in Chapter 8, we collaborated with two neighbourhood associations and a daycare centre, which was found via snowballing sampling. All the organisations lent us their venues to conduct the activities.

Working with populations in situations of vulnerability also required extra measures in terms of compensation, informed consent, and debriefing to balance respect and rigor [448, 348, 108]. Thus, compensation was adapted to the rules of NGOs, which required it not to

be excessive so that participants would not feel coerced. Extra attention was paid when it came to the requirements for informed consent: information was provided a month in advance, and the consent forms were explained to the participants out loud before the study started. Similarly, information for withdrawing participation was provided and displayed at the organisations' facilities.

Noteworthy, we adopted a different approach for RQ1. While the participatory principle of care remained, this study sampled a different type of participant: UX and UI professionals. We contacted professionals in the Greater Region of Luxembourg and the Netherlands using snowball sampling.

4.4 Data Analysis

This work aimed to open a conversation about the drivers that make users vulnerable to manipulative designs through their lived experiences. Therefore, it was not expected to cover exhaustively all aspects of vulnerability to manipulative practices. I cared here about subjective experiences rather than generalisable approaches [418, 264, 356]. With this epistemological commitment in mind, reflexive thematic analysis has been the preferred method to address the research questions.

Thematic analysis is a meaning-making process to identify patterns and trends in the data that allow us to understand specific phenomena. It is widely used in psychology, HCI, and design research because it allows researchers to create meaning around the patterns identified in the data, as well as to untangle nuances and deepen into contradictions and experiences [52, 51, 50]. Situating the experiences of manipulation into the idea of vulnerability calls for reflexivity and subjectivity as a resource [51, 161], even more so when adopting a harm-based approach. Thus, while participants talk about their experiences, it depends on the sensitivity of the researcher to understand and interpret when some of the experiences and harm come from the orchestration of design elements.

Quality criteria. I have carefully considered different methodological trade-offs to guarantee a robust and reliable method for all the individual studies presented in every chapter of this dissertation. Yet, I will give a general overview of the assessment of quality criteria in the context of qualitative studies building on [242, 423]: credibility, conformability, transferability, dependability, and reflexivity.

I ensured *credibility* of findings by triangulating data in different ways — that includes observation beyond what participants declared —, as well as contrasting negative cases within my analytical lenses. Data triangulation of different data entries, analytical lenses, and researchers has been specifically relevant to guarantee credibility as a quality criterion. Building on the literature on manipulative designs (See Chapter 3), it is unlikely that participants denominate and identify manipulative designs in the way the research community is labelling them. Hence, the level of interpretation and reflexivity of the researcher matters in linking the experiences of harm described by participants to the associated manipulative pattern. As detailed in my positionality statement (see below), my experience and position towards

manipulative designs have informed the interpretative analysis of the data — e.g. ‘How did my usage of the platform differ from participants?’ ‘why is the content I see different from theirs?’. To do so, I have selected a diversity of methods that complement each other in order to triangulate the data from different sources — e.g., observation in tasks, self-disclosure of felt manipulation, and perception of manipulative designs. The input from different data collection methods from the same participants aims to provide a more robust and nuanced understanding of their lived experiences [51, 356]. Similarly, iterative discussions with other researchers from different backgrounds helped me to fight potential biases that come from my own positionality [356]. For instance, for the study involving experienced designers in Chapter 5, collaborating with an experienced UX researcher working in Industry provided me with different lenses through which to look at the data. These collaborations with other researchers that combined different analytical lenses helped me to ensure *confirmability*.

I have worked on the transparency and thick descriptions of the results as a means to increase “*transferability judgment*”, understood as the extent to which other researchers could judge whether the results are potentially transferable to other contexts [423, 242]. In this regard, I have accompanied the results with verbatim and thick descriptions of why and how such verbatim happened, trying to find a balance between accuracy, detail and length of the explanations.

Given the nature of the selected methods, which looked at momentary experiences, it proved difficult to observe the results over time. Similarly, the specific logistic constraints from the collaboration with participants, made it challenging to member check after every study. Still, I aimed to ensure “*dependability*” as much as possible by conversing with the gatekeepers and workers of the organisations I worked with, since they had more extended experience with the participants. Lastly, I ensured “*reflexivity*” in every study by acknowledging my subjective position, embracing it as a resource [242], but also questioning it during the analysis of the data. I actively embraced my differences with the participants to confront my biases and assumptions and, therefore, challenge whether those would have an impact on the results.

4.5 Positionality Statement

The study of technology-perpetrated harms and the positions of vulnerability to manipulative designs has been my personal interest since the early stages of my career, even before engaging in an academic career. Thus, understanding how social inequalities may impact users interaction has been a primary motivator for the development of this thesis. It is important to be honest and transparent about where I position myself in this debate [161, 238, 51], which was born in interaction design but has several social implications.

“*Who am I in respect to my participants?*” was a recurrent question I asked myself along the journey of this work. I was a teenager whose parents could not provide me with the support needed to understand their digital surroundings. Although I cannot relate to the migratory experiences of my participants, who sought an opportunity in another place and whose families are not in the same country and; those have been my surroundings and my neighbours,

who inspired me in the development of the protocols. Seeing how the people surrounding me interacted with manipulative design has not only inspired me but also given me a closer understanding of some intimate relationships between the platform and them and how it actually impacts others and harms themselves. I will hopefully be an older adult who certainly feels the world is advancing too fast, and it is not my place anymore.

“How does my position separate me from my participants?” is even a more recurrent question I have asked myself, especially during the analysis of the different datasets. I have a strong personal position towards the topic, given my academic research and my commitment to a better digital world which already gives me specific lenses to look at platforms with manipulative designs. I have tried to immerse myself in the most common platforms that contain manipulative designs, believing that I could not do this research without knowing them personally. I was a frequent user of the video content platform Youtube Shorts and social media platform Twitter (now “X”), and I have become an even more frequent user of TikTok. Although I was already acquainted with several video games that use manipulative techniques, I have downloaded and played some of the most famous games to understand their simultaneous use of several manipulative techniques. I must admit that I found myself opening numbly Candy Crush several times during that process of platform discovery. My experiences with those platforms are definitely different from the ones of the participants in my studies. I normally enjoy discovering a new technique that some platforms use to nudge me to pay more or to give my personal data because I find it exciting for my research, and I am able to recognise them. I also receive very different content than my participants on those platforms — and I am aware of why that happens —, which has led me to ask so many questions in the analytical process. I embraced these differences as a resource during my analysis process [161, 51].

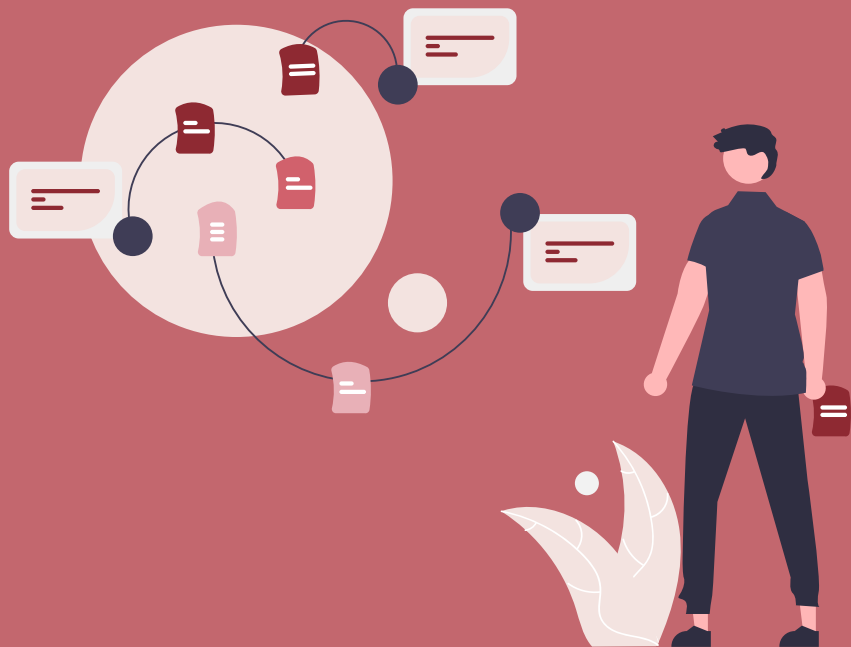
“Why do I care?” As stated in Chapter 2, the relationship between socioeconomic inequalities and interaction design exists, and the role of interaction design as a potential promoter or disincentive of inequalities seems crucial to me. My background in Law and Political Sciences has shaped my view about the existing tools to protect people, and has given me the lenses and the vocabulary to name them, and try to put them into practice. The knowledge of regulatory approaches, privacy, governance, technology, and justice that I have gained from my past background has conditioned my understanding of the topic and the directions taken in this dissertation.

Seeing myself, my friends, my neighbourhood, and my “reality” represented in the populations I studied, I could not help but intervene in the way I could: by doing research. Although I do not categorise this work as participatory design, it inspired me in some ways, and I believe in the words of Halskov and Hansen [186] when they call for the role of activist in participatory design research. Like so many critical scholars, I do not believe in design being agnostic, nor is research. Echoing Paulo Freire’s [152], “oppression is domesticating. To no longer be prey to its force, one must emerge from it and turn upon it. This can be done only by means of the praxis: reflection and action upon the world in order to transform it” (p.28). Hence, I believe that it is essential that once the oppressed are free from oppression, they have to combat oppression by practice. Otherwise, they will become the oppressor.

“How did I convey this to my participants?” When conducting the studies, I tried to ensure that participants knew I cared about the effects of technologies on users and that my mission was to help design better technologies for everyone. I never emphasized different situations of their potential vulnerability since I felt that could be condescending and stigmatising for them. Although I presented myself as a researcher, I also introduced myself as a local of the neighbourhood — for most of the participants —, I conducted the interviews in Spanish, and I tried to express care for their experiences and needs during my interactions with them [249, 435].

Part II

Conceptualising Manipulation in Practice



This part helps to conceptualise manipulative designs and vulnerability in practice for both designers and policymakers.

As part of the entangled actors that contribute to this problem and have responsibility over it, **Chapter 5** explores the tensions that designers experience when they are asked to influence users and how those tensions might contribute to vulnerability.

Following that study, **Chapter 6** presents a multidisciplinary conversation about the conceptualisation of vulnerability to manipulative designs. By understanding the flaws of legal texts in conceptualising vulnerability, the chapter presents the different ways in which users may become vulnerable, with a special focus on their ecologies, and provides some tools for legal scholars and policymakers to learn from HCI and design tools.

Ethical Tensions in UX Practice. The Fine Line Between Persua- sion and Manipulation in Interface Design

This chapter is based on the following peer-reviewed publication:

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. In Proceedings of the 2023 ACM Designing Interactive Systems Conference (DIS '23). Association for Computing Machinery, New York, NY, USA, 2408–2422. <https://doi.org/10.1145/3563657.3596013>

Abstract. HCI researchers are increasingly concerned about the prevalence of manipulative design strategies in user interfaces, commonly referred to as “dark patterns”. The line between manipulation and persuasion strategies is often blurred, leading to legal and ethical concerns. This paper examines the tension between persuasive UX practices and manipulative designs. UX/UI design professionals (n=22), split into eight focus groups, conducted design activities on two fictitious scenarios. We qualitatively analysed their discussions regarding strategies for influencing user behaviours and their underlying reasoning. Our findings reveal a combination of classical UI design strategies like sticky interfaces and incentives as their most common practice to influence user behaviour. We also unveil that trust, transparency, and user autonomy act as guiding principles for the professionals in assessing their ideas. However, a thorough approach is missing; despite a general user-first attitude, they feel constrained by contextual factors. We explain how the tensions between principles and context contribute to manipulative designs online.

5.1 Introduction

As a designer, how confident would you be to draw the line between a persuasive and a manipulative user interface design? A few decades ago, scholars and practitioners embraced persuasive design [276, 49], and captology [141], but would they be proud of all the resulting practices today [169]? A debate around when user experience (UX) practices become manipulation has been catalysed by the proliferation of tracking technologies [479, 428, 421]. Service providers collect more user information, leading to micro-personalised targeting, also known as “hypernudging” [260, 49]. Users’ autonomy is at stake: the extent to which this personalisation leads users unconsciously into manipulative practices online is questioned in a growing research community [343, 316, 297, 296, 49, 171].

The debate about interface designs influencing user behaviour has recently intensified among scholars, practitioners and policymakers. The term “dark patterns”¹ appeared in this breeding ground to describe “*user interface design choices that benefit an online service by coercing, steering, or deceiving users into making decisions that, if fully informed and capable of selecting alternatives, they might not make*” [296, p.2].

While the literature theoretically distinguishes coercion, deception, persuasion and manipulation in design [429, 48], it is hard to consistently tell them apart in the design outcome [380]. Persuasion convinces users with transparent arguments, while manipulation has a hidden influence on users [429]. Coercion implies giving users only one choice, while deception instils false beliefs by providing false information or hiding essential information. Scholars have discussed the threshold of when influencing user behaviour through design becomes an ethical concern in the context of soft paternalism, where users are supposed to be influenced for their own good [431, 432]. However, in the digital domain, the intentions of the service provider are often hidden, which makes it difficult for users to distinguish when they are being persuaded versus manipulated [429, 304].

It is urgent to study manipulation from the designers’ perspective because they can play a crucial role in addressing this problem [45, 374]. Design scholars have studied how social structures inform designers’ decisions [164, 83, 451, 23] and how designers can incorporate ethics in their design practice [153]. However, designers’ concrete approaches to designing for online influence have been understudied; while they are critical safeguards against e.g., manipulative designs. As Mathur et al. [297] explained, a crucial and open problem in regulating ‘dark patterns’ is understanding when online interfaces use admissible persuasive techniques without constraining user autonomy. To do so, practitioners need guidance to evaluate on which side of the persuasion line their designs fall. The presented study explores this gap: the tension between UX practices and manipulative design, by scrutinising the design strategies that can lead to user manipulation.

In the here-presented study, we investigate UX/UI designers’ strategies to influence user behaviour through their designs. 22 UX and UI professionals, split into eight focus groups,

¹This paper includes the term “dark patterns”, which is being discussed within the community to embed stereotyped connotations that impact communities of colour [143, 410]. The authors use it only for the sake of unpacking the problem of the ambiguity of this term in the community and propose alternatives in the outlook.

undertook a set of design activities on two ethically differentiated scenarios. We qualitatively analysed their discussions to identify their strategies to influence users and the principles and rationale that guide their design decisions.

Our contribution is threefold. First, we bring awareness to the grey areas when crossing the line between persuasion and manipulation, by unveiling how certain UX design strategies can become problematic. We, therefore, present strategies, principles and contextual factors that can facilitate manipulation, showing the tensions between them. Second, we provide empirical insights into practitioners' views on ethical designs when they seek to influence users. Third, we provide recommendations for designers to assess potential manipulation in their interfaces. Hence, this study provides a new lens to assess ethical practices in UX design, in the context of influencing user behaviours, showing how manipulation not only results from designers' intentions but also from their circumstances.

5.2 Related Work

5.2.1 The (Un)Definition of “Dark Patterns” as a Problem of Online Manipulation

Whether “dark patterns” is an ethical, psychological, or designerly term is unclear. While regulators have defined specific designs like pre-marked check boxes or providing inaccurate information about limited stock as “dark patterns” [387, 343], some policymakers [368, 56] refer to them generically as design artefacts in user interfaces that impair autonomy. This leaves room for interpretation when it comes to edge cases that resemble persuasive designs [111, 169]. In an attempt to define a “dark pattern” from a practical standpoint, Mathur et al. [297] distinguished attributes associated to the choice architecture of the interface. Gray et al. [169, 172] proposed categories of design strategies to differentiate the vast body of “dark patterns”: nagging, obstruction, sneaking, interface interference, forced action and social engineering. While both attempts to define dark patterns point towards questionable UX practices, some identified strategies overlap with classical persuasion techniques: for instance, emotional design, or computers as social actors [141] coincide with social engineering. This makes some “dark patterns” techniques hard to distinguish from persuasion [169, 49]. It is hence essential to further understand under which circumstances they become manipulative and cause harm to users.

5.2.2 Ethical Tensions Between Persuasive and Manipulative Design

In this paper, we discuss when UX design strategies fall under the realm of persuasion or when they become manipulation. We do not take a normative stance in deciding whether a certain manipulation is good or bad.

Scholars in applied ethics define manipulation as the intention to stealthily and elusively change a person's behaviour towards goals this person did not intend to achieve [331, 429,

376, 421]. For manipulation to occur, there must be an interest in subverting users' vulnerabilities covertly [376], typically with irresistible incentives. Manipulation reduces the user's capability of resisting the influence, and hence their autonomy [428, 304, 39, 233]. For instance, restricting choice options or the fact that they do not understand the influencing mechanism infringes their autonomy [39]. Manipulation may include deception and coercion [2, 428], although both are not necessarily manipulative. Deceiving implies making false statements deliberately to motivate a specific behaviour [428]. Coercion means leaving only one possible option to users, so they have no other choice [428], like removing options [39, 376, 428].

Unlike manipulation, persuasion does not infringe on users' autonomy nor exploit their vulnerabilities. Persuasion is a type of influence that is transparent about the persuaders' intent [105, 429, 142]. Trying to steer users in another direction does not imply unacceptable influence *per se* when it works transparently [105, 376]. This is what some authors call "rational persuasion" since the persuader gives rationales to the persuaded to motivate a different behaviour openly and transparently [39, 429]. Similarly, scholars have also discussed a specific type of influence: nudges [370, 459, 3, 304]. Depending on their transparency levels and resistibility, nudges can be manipulative or persuasive [337, 330, 304]. To be persuasive, nudges are supposed to be transparent or, at least, provide enough information to the persuaded and be resistible enough [304]. The differences between the types of manipulation and rational persuasion reside, therefore, in transparency and easy resistibility to the mechanism of influence.

5.2.3 How Does Design Influence Users? Persuading and Manipulating with UX and UI Design

How do we know when designs cross the line of manipulation, subverting vulnerabilities in a hidden way? Scholars in design have explained how, through the user interface design, affordances [160, 335] and signifiers [334] are made *apparent* and can apply a certain *strong force* [437]. According to Tromp et al. [437], the attributes of "salience" and "force" make designs decisive, coercive, seductive, or persuasive. "Salience" refers to how apparent the mechanism of influence is for the user. "Force" refers to the extent to which users have room to take a path different from the one that the design proposes [437], which translates into "user agency". To disentangle if something is a dark pattern -manipulative, hidden and subverting vulnerabilities- or simply persuasion -transparent and allowing users' agency-, designers have to navigate those attributes.

In an attempt to make the attributes "force" and "salience" navigable for designers, we have adapted the model from Tromp et al. [437] with applied ethics terms and theories (see Figure 5.1). Based on this adapted model, persuasion, manipulation, deception, and coercion can be represented as a function of salience and force. Persuasive designs are apparent in salience and weak in force since persuasion gives enough agency to the user in a transparent way. On the contrary, manipulation is hidden, not giving agency and subverting users' vulnerabilities, and can apply a strong or weak force. Coercion is a strong force, by limiting user

options, but it is also apparent to users.

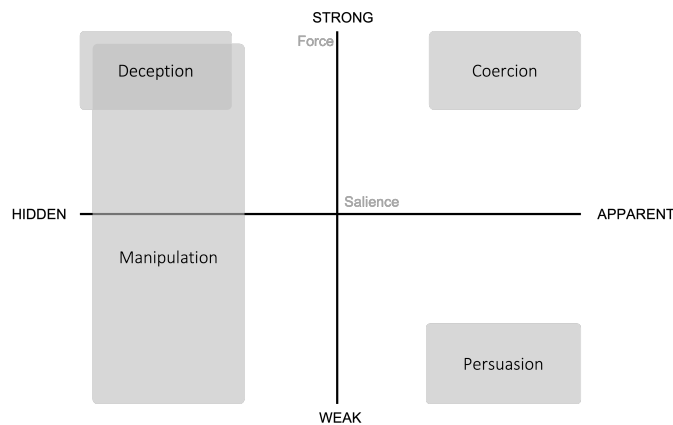


Figure 5.1: Dimensions of influence through design. Adapted from Tromp et al.'s [437]

Designers can take different ways to make design strategies more or less salient and strong. For example, friction might contribute to making a design strategy salient because it “interrupts the users’ primary tasks’ [118, 101, p.3]. Kollmer [239] explains digital sludges -i.e. nudges with bad intentions- through the concept of friction as a way to increase the effort for users to reach their initial goals. Scholars and policymakers have also criticised friction. For instance, Gray et al.'s [169, 169] taxonomy on designers’ strategies identifies obstruction and nagging techniques as dark pattern strategies that interrupt or redirect users. Bowles [49] criticised the use of salient persuasion as a potential solution that, in practice, may simply overwhelm users. However, friction in design is also used to foster reflection and guide users to more conscious decisions [101, 118]. Friction can even contribute to the force of a design by not only hindering but also helping users to resist the influence and being aware of their agency [281].

The use of emotions is a “hidden” or subtle technique that is widespread in UX design too. Fogg, explaining computers as social actors [141], already appealed to the use of human emotions in the same way as Norman, who coined the “emotional design” [333]. The “dark patterns” community is questioning the use of emotions as a way of manipulation since its resistibility is still unclear [169, 41, 343]. Without clear and universal criteria to set boundaries of how to use “salience” and “force” to influence behaviour ethically, we build upon the empirical exploration of designers’ rationales and strategies in UX to examine the tensions between UX practices and manipulation.

5.2.4 The Role of Designers in Designing Manipulation

How designers shape technology has been a concern of design and critical design scholars for many years [437, 460, 30, 6, 451, 22, 360, 206]. Shaping behaviours within the “digital architecture” [275], designers become mediators of ethics between business models and users,

implementing their ethical commitments [164] and being responsible for their impacts [30]. In their study with designers at their workplace, Gray and Chivukula [164] explained how designers mediate ethics in different ways. In some cases, they are aware of malpractices but constrained by organisational factors. In others, they mediate with their knowledge, applying and improving the design outcome and informing other stakeholders about problematic designs. Within this ethical mediation, Chivukula et al. [87] identified different roles that designers may adopt when applying ethics: educator, learner, policy-follower, activist-advocate, member of my profession, responsible, deliberative/thoughtful and translator. Chivukula et al. [83] also noticed how design students leveraged different values according to how they interpreted the business intentions: donations for charity versus manipulative e-commerce. User-centred approaches and persuasive techniques sometimes conflict in the design process, confronting values between designers and other stakeholders.

If designers do not mediate on ethics, contesting specific stakeholders' decisions, there is a risk of falling into "ethical blindness" [350]. "Ethical blindness" refers to acting unintentionally unethically in certain circumstances. For instance, when stakeholders have a fixed set of values that nobody contests [350]. In this regard, Wong [462] explained how UX designers used tactics of soft resistance to confront different values of stakeholders when designing technology. The relationship between stakeholders plays a crucial role in some ethical frameworks [153], but also is an essential part of user-centred design and UX approaches [256]. "Ethical blindness" is present in anti-patterns or the #asshole design discussion [166, 135] as bad designs that are not necessarily intentional but lead users into decisions they would not want to make. We expect to contribute to this set of literature explaining these tensions for professional design practitioners in the specific context of manipulation in design.

5.2.5 Research Questions

There remains a gap in prior research about a clear differentiation between persuasive UX practices and manipulative designs. The present study will address this gap. Our objective is to investigate when UX design strategies cross the line of manipulation by understanding design practitioners. Therefore, we empirically study their design strategies employed to influence behaviour, their considerations towards those, and the context in which those design practices take place. To reach those objectives, the present study aims to answer the following research questions:

(RQ1) What strategies do UX designers use when they are asked to influence users through an online interface?

(RQ2) What are UX designers' considerations when evaluating their design ideas to influence user behaviour through an online interface?

(RQ3) What are the contextual factors that shape UX designers' decisions when they are asked to influence users through an online interface?

In studying "dark patterns" and manipulative designs, designers have not been adequately involved in disentangling these practical aspects. Although Gray et al. [169] use a practice-led approach, we still lack insights from design practitioners into their persuasion strategies

and their pain points when designing influence and assessing the effectiveness and ethical boundaries of their designs.

5.3 Methodology

5.3.1 Overview

The three research questions were addressed through a qualitative data collection based on a series of design activities. UX/UI design practitioners were asked to execute a set of design tasks on a realistic brief. We conducted a thematic analysis [52, 51, 50] of the conversations with UX/UI design practitioners using inductive coding to uncover their approaches and the tensions within the design strategies when influencing users.

5.3.2 Participants and Recruitment

We recruited 22 UX/UI professionals currently active in industry, with a wide variety of backgrounds and years of experience. They had on average 8.35 years of professional experience (mode = 9, sd = 7.7). 15 identified themselves as women, 6 as men, and 1 as non-binary (see Table 5.1). We used professional networks and social media to reach design practitioners: potential candidates whose LinkedIn profile included "UX/UI designer, manager, consultant" or similar terms were contacted. Then, the contacted participants were asked to invite their colleagues to participate in the activity, as a means of snowballing. The participants were distributed into eight groups of 2-3 participants each and assigned to one of two design briefs: Four groups conducted the assignment to design for a human rights NGO (from now on "NGO") and four groups for a fast-fashion (from now on "FF") website (see Section 5.3.3 for more details). Seven of the groups were formed by co-workers who respectively worked for the same company, meaning seven different companies were involved. For one company, we formed two groups for the design activities (company B in the table below). The remaining group was formed by designers who did not know each other previously (company G, H and I). The study was approved by the Ethics Review Panel of the University of Luxembourg and complied with all relevant regulations. Informed consent was provided by participants, who were fairly compensated for their time.

5.3.3 Protocol

Activities

Each group of participants was given with a design brief simulating a real task. In both briefs, they were asked *"to create an effective solution for [the client's] website to ask visitors for their emails to receive a newsletter"*, similar to the scenarios proposed by Chivukula et al. [83]. Depending on the brief, participants drafted a solution for a "Human Rights NGO" or a "Fast

Table 5.1: Distribution of participants by company, brief, job position and years of experience, educational background as self-declared. FF corresponds to the “Fast-fashion” brief, NGO corresponds to the “Non-Governmental Organisation” brief.

Comp.	ID	Brief	Job position	Exp.	Education
A	P01	NGO	Senior UX Designer	9 years	Electrical and Computer Engineering
A	P03	NGO	UX Designer	8 years	Engineering in Computer Science / Design for Interaction / Industrial Design Engineering
A	P19	NGO	Senior UX Designer	11 years	Computer Engineer Design / Fine-arts / User-Centered Design
B2	P08	FF	UX Designer	5 years	Design
B1	P09	FF	UX Designer	10 years	Web technologies and management
B2	P10	FF	UX Designer	30 years	Computer and Accounting (Development)
B2	P12	FF	UX Consultant	3 years	Communication and information
B1	P04	FF	UX Manager	9 years	Communication and Computer Graphics
C	P11	FF	UX Designer	4 years	Information and Communication / Digital Project Development
C	P13	FF	Lead UX/UI Designer	9 years	Graphic Design / Web Design
D	P05	NGO	Senior UX Designer	6 years	Psychology
D	P14	NGO	Fronted Developer	4,5 years	Engineering in Computer Science
D	P15	NGO	Product Owner	13 years	Business / Innovation Digital Management
E	P16	NGO	UX Designer	3 years	Product Design / Advertising / Digital design and project management
E	P17	NGO	UX Designer	8 years	UX Designer / Computational Sciences
E	P18	NGO	Head of Design	29 years	Modern Design
F	P20	FF	UI/UX designer	3 years	Architecture
F	P21	FF	UX/UI Designer	2 months	UX/UI Design
F	P22	FF	UX Researcher	3 months	Cognitive Science
G	P02	NGO	UX/UI Designer	2 years	Computer Applications
H	P06	NGO	Freelance Designer	10 years	Cognitive Science / Computer Science / Socio-technical systems
I	P07	NGO	UX/UI Designer	7 years	Videogames Development and virtual environment

Fashion company” that needs to collect email addresses to increase awareness and participation in human rights activities or to send commercial offers in the case of the fast-fashion company. With the exception of the business’ intention (NGO vs FF), the instructions and tasks provided were the same in both briefs. We chose these briefs because while they represent different intentions from the business’s point of view, neither business must necessarily have their intentions aligned with the user. Solutions were sketched as low-fidelity hand-drawn prototypes (from now on ‘the prototypes’). The participants were informed that they could be creative, check information online and write comments on their designs. The activities were conducted in a university lab or in premises provided by the companies respectively.

The session was divided into three activities:

- i In the first, individual, activity, the participants were asked to design an effective solution for their scenario. To stimulate ideation they had to make a Crazy 8 activity [21] (which consists of sketching 8 solutions in 8 minutes), before further detailing their preferred idea (20 min). Participants then pitched their design idea to the group (2 min).
- ii This was followed by a group activity in which participants were asked to design one solution together, either based on their individual ideas or from scratch (20 min). The group then pitched the solution to the moderator (2 min).
- iii Lastly, a focus group was conducted to reflect on the prototypes and the strategies to influence behaviours online (30 min). Only in one session, with designers from company B, we conducted two parallel design activities, combining both sub-groups for the focus group at the end of the session. In the focus group participants were asked about the following topics: elements that influence users, users’ understanding of influence, the impact of design elements on users and society, designer’s perception of ethics and designer’s perception of dark patterns.

Material

Participants were provided with the following material: (i) Design brief, (ii) a Crazy 8 template [21], (iii) a layout with the landing web page of the FF or NGO website made by the researchers and inspired by actual websites. (iv) A collection of royalty-free images corresponding to the topic.² (iv) Templates to elaborate their design ideas.

5.3.4 Data Analysis

We recorded and transcribed the participants’ individual and group pitches, as well as the focus group discussions. Given the topic’s complexity, and its subject-dependency, we built our analysis on the reflective thematic analysis method (from now on “RTA”) [52, 50]. We looked for patterns in the participants’ approach to influencing user behaviour and the tensions they

²Licence CC- BY 4.0. Authors of the images. Calicadoo, Karolina Grabowska, Luobulinka, Artem Beiliiaikin, Alyssa Strohmman, Cotton Bro, Dom Hill, Moose photos, Doug Linstedt, Ayo Ogunseinde, Rosemary Ketchum, Koshu Kunii, Jordy Meow, Paddy O’Sullivan, Moa Alexanderson

Table 5.2: Themes identified in the data and their alignment with the research questions

Research question	Overarching themes	Themes
RQ1: What strategies do UX designers use when they are asked to influence users through an online interface?	Influencing is an exchange	Convincing users with reasons Exchange between users and designers Understanding how strategies work on users Friction and stickiness
RQ2: What are UX designers' considerations when evaluating their design ideas to influence user behaviour through an online interface?	Conditions of manipulation and guiding principles	Conditions of manipulation Impacts Autonomy Trust and transparency Usability Users vs user experience first
RQ3: What are the contextual factors that shape UX designers' decisions when they are asked to influence users through an online interface?	Responsibilities and hurdles	Design responsibility is shared Imbalanced power towards business

experience. For the sake of reflecting on the verbatims and understanding transcriptions, we must note that there was only one participant whose mother tongue was English. The verbatims accompanying the results are therefore corrected.

We built on the Braun and Clarke [52, 50] six-stage process of reflexive thematic analysis to actively seek the themes and sub-themes. The first author, who conducted the data collection and transcribed the data, familiarised themselves with the data during the transcription process, by extracting summaries of the data and taking notes regarding relevant insights. Then, initial codes were generated. The software MAXQDA supported this process. Alternating the scenarios while searching for themes allowed to uncover more nuances and evolve from more descriptive codes to latent themes. Relationships between codes, sub-themes, and themes were established at this stage, supported by visual maps. These themes were confronted and discussed with the second author, who double-coded 50% of the data. Themes were reviewed and data was reinterpreted through discussions in an iterative process. Finally, the three overarching themes that answer the research questions were defined and named, with a sub-level of themes (See Table 5.2).

5.4 Findings

In this section, we present the analysis of the discussions with the participants, grouped by each research question. First, we present the strategies that participants use when they are asked to influence users through an online interface (RQ1). Second, as designers' considerations (RQ2), we present participants' delimitation of manipulation and four main guiding principles they apply when evaluating this in their designs. Third, we report the contextual factors that shape participants' decisions when they are asked to influence users through an online interface (RQ3).

5.4.1 (RQ1) Design Strategies to Influence Users: Influencing is an Exchange

"Influence is an exchange" captures the idea that participants perceive the process of influencing users as a trade: users cannot be asked something without obtaining something in return. This idea comes from the principle of "*good user experience first*" (see Section 5.4.2) and informs the strategies designers use to influence users through their designs. In response to RQ1 we elicited the following strategies: convincing with arguments, providing tangible and intangible incentives in exchange, using frictional and sticky elements and testing the effectiveness of the mechanism of influence.

Convincing users with reasons related to the product/service

The main strategy participants adopted to convince users was outlining the direct benefits or favourable consequences. This strategy was especially adopted in the NGO scenario, in which participants pointed to the NGO's positive role in the world to convince users to sign up for the newsletter. The FF participants used rational arguments associated with the products and what the users can get through the products.

Exchange between users and designers

Apart from highlighting existing benefits, participants also tried to introduce new, additional benefits to the exchange. Advocating for giving a pleasant user experience, the participants always tried to provide added value when they asked for users' email addresses. Even when the brief was not offering anything, they would attract the user with a presumed added value - e.g., special offers for your birthday (B1). The participants fear that users might consider the newsletter a nuisance instead of something with potentially inherent value. Participants provided this type of transaction to users by three principal means: *appealing through feelings*, *engaging with the audience*, and *offering tangible and intangible incentives*.

Participants showed reluctance to use negative *feelings* while using positive ones was perceived as a common way of providing a pleasant experience. This is a strategy to give more weight to the positive side of sharing the email address, positioning it as an *intangible incentive*.

Some intangible incentives offered were feelings: belonging (D), community (A, B1, B2), hope (E, A), trust (A, F, G, I), welcome and connection (B1, B2). On the other hand, *tangible incentives* such as stickers (D), platforms and podcasts (A), money (B1, B2, C), or even a styling service (F) were also offered through the prototyped solutions.

P5, P14, and P15 illustrated the different types of incentives they provide to engage users. They initially suggested a physical reward such as a sticker, ultimately leading to the intangible reward of recognition for being part of the NGO. These types of exchanges were sometimes identified with the idea of “user first” (P8) as a way to provide “meaningful experiences” (P2). Still, they are intentionally designed to influence users.

M: *“I’m very curious because this is not the first time that the three of you mentioned transactions or incentives. What do you mean by that in this case?”*

P15: *“The stickers. Right? Yeah.”*

P14: *“She had the idea of stickers, but it depends if you can give more incentives...”*

P5: *“It’s not like stickers. It’s more than just the stickers, right? It’s also like, if you put the sticker, like on your car or on your bag, there’s also like this feeling of ‘I’m a good person because’... and that’s free marketing for the company.”* (See Figure 5.2)

Friction and stickiness

Friction and stickiness represent concrete UI mechanisms that participants balanced to influence users. Friction is a way of interrupting users’ tasks, and stickiness is a way of repeating the call to action, so it always stays present in the user’s sight. Building on the “user experience first” principle, participants were unanimous in cautiously using friction to catch the user’s attention without bothering the user. Hence, we observed the trend of using sticky mechanisms to keep the information present as long as possible in the prototype, without being frictional, but always prominently positioned in the hierarchy of the prototype. These mechanisms were observed for the use of pop-ups, sidebars, top bars and calls to action. P10 explained how keeping sticky calls to action, which are always visible when browsing through the website, will continuously catch users’ attention to subscribe.

P10: *“And the way you can see, because you have, for example, the pop-up once. But if you don’t respond now, you have a possibility with a fixed banner to [come back] when[ever] you want [...]. And the other one is, for example, you don’t need to subscribe [now], but once you go to the basket, you have a reminder because you already [gave] your email address. So you just have to uncheck and check, [there are] several positions [to place the call to action]... it could be interesting.”*

Understanding how strategies work on users

To conduct an exchange with the user, participants reported the need to understand how their prototypes work on their actual target. They suggested the standard techniques of testing interfaces like A/B testing (G, H, I, C, A), and personas (P15). They would also benefit from

using analytics (B1, B2) from the website and from third parties. While participants expressed their desire to better understand the users: *“talk a lot, test a lot”* (P21), they actually mentioned methods that ultimately aim to exploit the users’ weaknesses while guaranteeing a good experience. This shows their desire to understand the users stems from the intention to elicit a certain behaviour out of them.

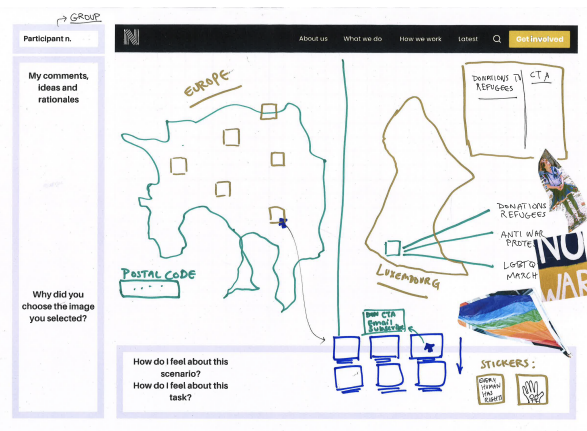


Figure 5.2: Prototype from group D for the NGO brief. Participants decided to maintain the call to action persistent with ‘windows’ that blink showing different parts of the world. The intention is to avoid an intrusive design for the user.

5.4.2 (RQ2) Designer’s Considerations when Influencing Users: Delimitation of Manipulation and Guiding Principles

During the design activity, the participants regularly self-assessed their solutions [19]: Is this design going to be effective in obtaining users’ email addresses? How can I make it even more effective? And how would I react if I was facing this interface? In response to RQ2, we elicited participants’ considerations when evaluating their designs based on two aspects: their delimitation of manipulation and their intrinsic guiding principles when designing.

The limits of manipulation

As a general trend in our data, participants demonstrate intuition about what they deem acceptable influence and what crosses the line, but the concrete limit of manipulation is unclear to them. The researchers did not introduce the term manipulation; it spontaneously emerged in the participants’ conversations and proved very difficult for them to define clearly. Manipulation bore a negative connotation throughout our participants’ exchanges. They associated manipulation with restricting user autonomy, “pushing” users into something unsolicited, or tricking and playing with them. However, reflecting on what is socially acceptable in offline marketing and television, the participants found that their techniques resemble those and

are hence acceptable: *“so that’s the beginning of manipulation, but that’s just the way it is, that’s communication”* (P16).

Role of the client. The nature of the client seemed to trigger ethical concerns for participants. Except for one group with the NGO brief (D), none of the designers reflected if they were manipulating, associating and justifying ethical responsibility with the business behind the request *per se*: *“in my opinion, for the NGO or another governmental institution which tries to help people and tries to save the planet [...] [if] I must design a mega massive button for [the user to] subscribe, in my opinion, it is not a dark pattern”*(P18), or *“the shop, the fast-fashion is not ethical, it is the company”* (P4). As an illustration of this trend, one significant difference between the NGO and the FF briefs was the use of “real stories” (D, A, E). Participants with the NGO brief would not consider their strategies manipulative if the message explained a reality within the organisation - e.g., projects to prevent war, hunger or poverty. On the contrary, the participants on the FF brief would acknowledge their potential manipulative power when using similar strategies.

Impacts of design. We expected participants to also reflect on the potential unintentional consequences of the design choice during their exchange. Only two prototypes embedded design decisions with inclusion (B1) and ecological purposes (E). However, during the focus groups, the moderator had to explicitly ask about the impacts the prototypes might have because they were rarely spontaneously addressed by the participants. Despite this facilitation, the participants only reflected on their intended outcomes -i.e., increasing subscribers- as a potential consequence, while it was harder for them to foresee unintended consequences.

Participants were then also prompted to talk about the user groups and the societal aspects their designs might affect. When asked about the impact of their design on the user, the answers still did not go beyond the intended outcome or the direct effect of the design on the target users. The participants barely saw any potentially negative impact, with few exceptions: privacy loss and the vulnerability of users with high sensitivity to Human Rights when using images that appealed to emotions (D, E). When asked about the societal impact, participants seemed to be aware of some potential unintentional consequences caused by their design outcomes, such as an excessive carbon footprint or a lack of diversity. To counter such consequences, participants considered green design solutions (C, A, E) and the use of inclusive images (B1, B2, E, F).

When discussing the potential susceptibility of users towards the design strategies, participants showed a pattern of identifying inter-sectional conditions of vulnerability: personal sensitivity towards certain topics - users more connected to the topic, like the LGBTQ+ community (D, F), but also lonely or isolated people (B1, B2) - and socio-economic vulnerabilities - young people, people with fewer resources, or women (B1, B2, C). Especially in the FF brief, the focus group reflected on potential harm for people at the intersection of youth, low economic status and personal vulnerability. *“The problem with fast-fashion, is [for] people who have less money. It’s not always young people, but people with less... buying power”* (P4).

Guiding principles

The main guiding principles that the participants drew on to avoid manipulative designs were the following:

Autonomy. From the UI elements and general strategies of influence proposed by the participants, we saw that they aimed to increase trust to ensure that the users agree “voluntarily”. Thus, coercion and deception were perceived as unethical influences that participants would not apply. While they expressed that leaving the decision to the user is crucial, they had no clear idea of how to give such agency to them. The participants argued that giving agency to users relates to the amount of information shared with them: if users are informed, this will allow them to make rational decisions.

P15: *“And if you’re gonna give me your information, of course, make sure that [you] are agreeing on receiving the newsletter and agreeing on giving someone else’s email. And also, this [other] person has to agree to be subscribed. And if we had all these steps, yes, I guess it’s transparent.”*

P15 alludes to the fact that with enough information, regardless of the mechanisms that are being used, users always have agency because they can close the webpage. In relation to this, P11 explains, the website gives autonomy because *“[it] gives you the freedom to make your action or not. Here you are free not to give your email address, but you have 20 per cent off”*. Users are autonomous to actively avoid the influence: rejecting the discount, avoiding the pop-up or closing the website. This alludes to an interpretation of autonomy as negative liberty: as freedom from external barriers, instead of a positive one, freedom to act [73]. To act upon this negative liberty, participants assume action on the part of the user - e.g., closing the window.

Autonomy as a core principle is also present when participants are asked about “dark patterns”. They easily identified those as strategies to *“trick users”* (P9) through forcing, deception, manipulation and coercion, and they expressed that they would never want to find themselves consciously designing such elements.

Trust and transparency. Participants associated autonomy with transparency and trust, as this would guarantee users’ autonomy. According to the participants, users need to trust the client and their service in order to be influenced. To achieve this, the participants aim to create designs that are transparent with the user about what is being asked of them and what they obtain in exchange. Hence, this was reflected in the prototypes using contact information, privacy policies, or more information links.

P7: *“[The importance of trust is] for the user, who needs to click on the button. Because if you’re not sure what is going on with either the NGO thing or to whom you are giving the information, maybe it’s another entity behind it managing [the data]. [You] don’t know if it’s a mail service that will send you a lot of spam [...] or [if] you get scammed [...] So here we should need to work more on the trust and of course [this] is the first step to building a website to get the [user’s] email. But yeah, the user will need more trust on what we are providing.”*

P7 and P2 discussed the need to disclose as much information as possible to instil trust in the user and therefore get users' emails. This reflects a nuanced conception of transparency as a value, as it could be considered instrumental to the goal of influencing the user, instead of a value per se. There is a second nuance in the idea of *transparency*. It is not clear if participants refer to the transparency of the business intentions or the transparency of the influence strategy. Instead, the transparency of the strategies depends on participants' preconceived notions of what users already know about these strategies, similar to conventions regarding usability. As P11 explained: *"sometimes it's better to make something that is pretty similar rather than very 'innovative', because [you] lose the user or create some confusion"* (P11). This connects with the principle of *usability*.

Usability. Participants considered *usability* as another guiding principle. They reflect on users' expectations for the specific use case and then rely on the design strategies and mechanisms that users are familiar with, to make the interaction as smooth as possible. Although participants recognised that they might be excluding specific target groups -like the elderly or users with disabilities-, they generally employ common elements, such as pop-ups, buttons to close, banners or sliders. They furthermore favour design patterns that have proven effective for other providers: *"The things that we saw really often [...] [are] the ones that I'm sure will work for 20 or 30 per cent of the users"*(P13).

Users first vs good user experience first. Participants identified themselves as advocates for users; however, they are actually advocating for a good *user experience*. Participants chose their design strategies considering users' needs, users' expectations, and, above all, the user experience. This distinction is important because guaranteeing a good user experience implies that practitioners disregard strategies that can be bothering, annoying, too pushy or intrusive for the user. However, as a counterpart, they may create more subtle mechanisms that guarantee a good experience but with the power to subvert vulnerabilities. This is very present in the use of feelings as strategies (see Section 5.4.1).

P3: *"[...] at the same time, [I] have like a message of hope. So, when I look at all these images (Researcher comment: Participant points out some images) they're depressing and... which is true. Like... it is a very strong issue, and it is something that people should feel strongly about. But oftentimes, [the users] need inspiration as well. [They] need a place of hope, which is why I chose images that are sad, but also they look up to something positive so which is why there's this... yeah, this girl [among] ruins."* (See Figure 5.3)

Participants did not mean to trigger unpleasant experiences with, for example, shocking images in the NGO scenario. They aimed to evoke positive feelings as a way of getting the user to subscribe to the newsletter. This is also present when testing mechanisms (See Section 5.4.1): participants' desire to test stems from the intention to change behaviour. The participants, therefore, knew that a good user experience is instrumental to influencing user behaviour.

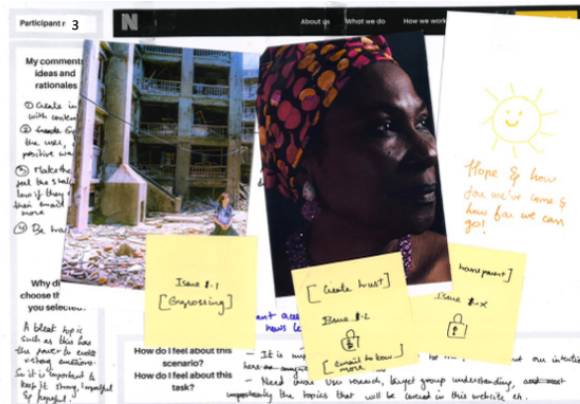


Figure 5.3: Participant 3's prototype for an NGO. The participant selected three different images to transmit hope, transparency and trust to the users. This was a matter of showing what the NGO does and providing users with intangible incentives that nudge them to provide their email.

5.4.3 (RQ3) What are the Contextual Factors? Responsibility and Hurdles in Designing Influence

To answer RQ3, we found the context in which design practice takes place contributes to determining the use of online influence strategies. Designing ethically is a shared responsibility and depends on the support or constraints coming from other stakeholders. However, the position of designers in this decision seems imbalanced. Therefore, the set of stakeholders with decision-making power, and the position of the designers within the organisation, matters.

Responsibility is shared

Participants reported being aware of their responsibility and their knowledge about human psychology, being able to manipulate users. Their role is crucial, but they are not alone in this process: the set of stakeholders with decision-making power matters. This was shown when reflecting on other stakeholders as supporting or constraining factors in the design of influence. Participants regularly mentioned the reliance on the role of the legal department and corporate values to ensure that regulatory limits are not trespassed by their designs. For example, they know about GDPR compliance but are not necessarily sure about its implications for their designs. Web developers are also referred to as key collaborators: the feasibility and cost of implementation for web developers were considered when assessing the suitability of the solutions.

Participants know that the client plays a significant role in how ethical their actions can be; hence, they sometimes associate ethical responsibility with the business. While FF participants tend to identify unethical practices in malicious designs required by the business model, NGO participants justified their potentially manipulative designs with the 'greater good' that the

NGO would bring to society. However, participants failed to recognise other impacts from their design that are not associated directly with the business.

Participants unanimously agreed that the contracting client/main stakeholder holds the responsibility for the design choices: *"We use emotions for the client says [inaudible]. It works, but it's not very grateful"* (P12). There is a clash of values between what designers consider permissible and the business interests, which is further developed in the following section (see Section 5.4.3).

Imbalanced power towards business

This set of relationships constrains participants and their power of action when it comes to behaving more ethically. They acknowledged they have two options to avoid manipulative designs: carefully choosing the company they work for and rejecting the assignment while assuming the consequences this implies.

P9: *"[...] we are working for [this company], personally, as an internal designer rather than working [...] for Amazon, for example. It's a choice because I don't want to help Amazon. They're making a lot of money. I don't care about Jeff Bezos. So I prefer to work here and help the poor [employees of this company]."*

Similarly, they advocated for educating other stakeholders in the company in two ways: firstly the best ways to increase subscriptions at any cost, and second how to do the same without intentionally harming users. Here, P12 and P9 explained their educator role. They can teach that avoiding unethical practices, despite the sacrifice of some subscriptions, can be worthwhile.

P12: *"I agree with with [P9]. I think it's good to also show [to future colleagues] what it is, what exists, like dark patterns, and maybe give them some solutions to do things well. And also I agree because I think it's also our responsibility to maybe, maybe not change because it's going to be complicated; but maybe to be conscious about things. And, for example, like for the newsletter, I think maybe we can explain to the client that, sometimes, it's better to have maybe less number of emails but with more quality. It's better to have maybe 20 than 100."*

However, some participants acknowledged that acting according to or imposing their values is not always possible, and their role as educators cannot always be fulfilled. P6 and P7, who were freelancers, discussed their actual possibilities of avoiding the design of "dark patterns" if the company asked for it, explaining the difficulties of their position. The designer's reputation and values might be at stake, but it is not an easy decision when "you need to eat" (P7). P6 and P7 exemplify the internal conflict that designers can face when trying to be ethical: they have their own values, but they are afraid of finding themselves rejecting everything that might look unethical.

P6: *"I think, personally, I will feel the spirit before signing any contract. But we never know if it's something really big. Maybe I will ask myself as a designer if I want to be*

associated with that. That would be subject to one contract because I work for myself. So, if I want to show my work, I have to show my values too. So, I don't want to reject every touchy subject... Mhm...maybe I will try to find a way to engage myself as a designer for that kind of thing and say okay that part I don't do it and you will find someone else."

P7: "Well, sometimes you have the option to say no, I don't want to design something like that because I don't agree with this position of the company/organisation [...] So as you say, [P6], it is a bit hard and we need to eat, we need to [earn a] salary, and you need to pay bills and so on. So, in the end, you say 'Okay what [am I] doing here? I'm doing business'. So, in the end, you need to provide the business value and you need to accept some stuff. But the recommendation is to don't do that, to just avoid it."

5.5 Discussion

Through the analysis of the discussions, we found that while participants' guiding principles seem to be a driver in implementing UX strategies, there are various reasons why they fail to implement them properly. This mismatch between guiding principles and design strategies seems to be the opportunity for an unconscious design of manipulative strategies. In this section, we want to highlight the following tensions between UX practice and ethics. First, we will explain the tensions between the designers' guiding principles and the contextual issues. Second, we will discuss the strategies of UX practice that might lead to manipulation in the absence of a proper ethical analysis, and we will consider its tensions with the designers' values.

5.5.1 Tension Between Principles and Contextual Factors: An Imbalanced Ethical Mediation in UX Design

We have highlighted several contextual factors that put designers in complicated positions, as these limit their decisions. These results are aligned with Gray and Chivukula's framework of ethical mediation in design [164]. The authors explained how designers mediate ethics in three ways: they are constrained by organisational practices, they impose their individual practices in the organisation, and they use theories from applied ethics scholarship in their design. The insights from our focus groups show how designers sometimes impose their individual practices over organisational ones: choosing who they want to work for and instructing stakeholders in ethical practices. Our results also highlight the opposite situation: organisations impose their own agenda on designers, making them accept their limitations as a designer and subsume into stakeholders' practices. Participants reflected on how accepting or rejecting work from their client or employer sometimes is their only way to make ethical design decisions. Our results do not bear much information on how applied ethics shape design practices. Only the applicability of data protection regulation as "applied ethics"

seemed to shape their practices. Within this set of relationships, we have found how participants identified themselves with different roles explained by Chivukula et al. [84], like educator and advocate, but with a clash of identities that is difficult to resolve. It is noteworthy that our participants had difficulties when reflecting on the potential impacts of their designs. This contrasts with the reasoning of Wong's [462] participants, who use soft resistance techniques that try to address potential impacts on underrepresented communities. It is, therefore, necessary to reflect on these results together. Designers might not be trained for the assessment of impacts. However, as shown in our results, they also rely on other roles which hold, in Wong's terms [462], the "ethics ownership": business, managerial or legal departments. This raises several questions: to what extent is ethical mediation possible under certain circumstances? To what extent can we make designers responsible for ethical assessments when they are immersed in an ecosystem of actors? Reflecting on this matter might give us an insight into the problem of 'dark patterns' in large digital services.

Designers are not moral philosophers, and, although they have an intuition, we cannot always expect them to make a complete assessment of the ethical aspects of designs. Without the proper tools and knowledge, the way they mediate ethics becomes imbalanced (See Figure 5.4). Applying the framework of ethical mediation to our sampling of designers, we see an imbalanced set of relationships in which applied ethics is limited, weighing the organisational practices much more than individual practices. Although we agree with the scholars who advocate for a better integration of ethics in education and design programmes [413, 156], there is also a need to rethink the whole governance system of technological design.

Designers need a better system of checks and balances within their companies that allows them to provide their expertise and ethical commitments without the risks of suffering adverse consequences. Untried mechanisms of balancing ethics in design are being discussed by policymakers, internet governance and applied ethics scholars. They include ethical impact assessments [291], documentation of the design process [41, 343] or self-regulatory instruments [284] — e.g., the use of codes of conduct for UX design in the organisation. These instruments could help designers reflect on their designs' impacts, which our results have shown to be problematic. Providing these tools to designers might contribute to facing "*the imbalanced power of business*" and could include all the stakeholders that advocate for users and good practices. Moreover, including legal departments, DPO (Data Protection Officer) or ethics departments within those governance systems will empower designers to implement proper "applied ethics" with other stakeholders' inputs.

5.5.2 The Tensions Between Principles and UX Strategies: When Persuasion Becomes Manipulation

Aside from the clash of principles between stakeholders and designers, the way designers interpret their guiding principles and how they implement their strategies gives rise to potentially manipulative designs. This is especially aggravated by the context of the surveillance economy [479, 49, 260]. Designers believe they are advocating for the user when they are ac-

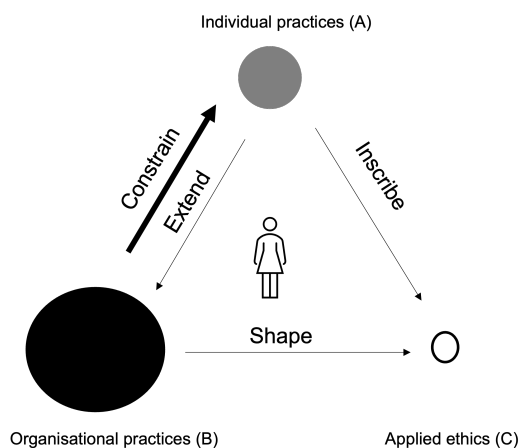


Figure 5.4: Gray and Chivukula's [164] framework of ethical mediation applied to our participants. Adaptation from the original.

tually conducting user profiling, “hypernudging”, and opening the opportunity to subvert the user’s vulnerabilities and weaknesses, in other words, to manipulate.

Tensions in autonomy: creating irresistible incentives in a hidden way

The implementation of ethical principles is a double-edged sword. The way designers preserve user *autonomy* through the designs is the main point to assess whether a particular design includes manipulation [429]. Our results highlight two types of UX strategies that, if used to target user vulnerabilities, may become irresistible for users. Using *tangible incentives*, specifically money -or discounts- seems irresistible enough to become manipulative. Using *emotional triggers* about sensitive topics for users might be irresistible in some cases. What can be perceived as a rational way of persuasion can easily move to the realm of manipulation if the design strategy works subtly. Providing convincing arguments about why you should give your email address to a fast-fashion website, showing the provider’s intentions, falls under legitimate persuasion; when those arguments subvert vulnerabilities in a hidden way, however, they turn manipulative. *User research and analytics* methods to test the strategies on the users, constitute a perfect tool to investigate the exact points at which users misclick, misread the information or are more inclined to specific topics. Designers know these methods work; hence, they incorporate them into their toolset.

Tensions in usability: when the “average” user becomes the “vulnerable” one.

Usability as a principle also might increase tensions. Designers adjust their designs to standards that the user might expect, and the user, in turn, gets accustomed to the designs they interact with. UX methods foster the focus on user groups with their specific needs creating a sort of “average user” for their designs; this might be problematic in the realm of influencing

behaviours. When UX designers are aware of psychological and socio-economic vulnerabilities that make their designs more effective for a given group of users, they might unconsciously subvert those vulnerabilities. Reflecting on all users and what characteristics may render some of them more vulnerable, should be an integral part of every design process to prevent from unintended exploitation of vulnerabilities.

Tensions in UI elements: how UI design contributes to a misimplementation of principles.

The interplay between values for the design of UI elements might also create irresistible strategies. Although *transparency* and *trust* are essential for designers, the interpretations of those may lead to malpractices. Providing information about the provider is a way to generate trust. Yet, if the only purpose is to attract users, then transparency is not an ethical commitment, and it may become a manipulative strategy. The consequences of value leveraging are reflected in the idea of *friction and stickiness* as mechanisms. Despite our participants' conviction of not forcing or coercing the user, there is a lack of reflection on the idea of resistibility to the influence, which is key [304, 429]. Designers, trying to guarantee *positive user experience first* -avoiding "bothering" frictional calls to action-, make more subtle designs that, with the wrong combination of elements, might become manipulative. Frictional elements, therefore, need to be analysed in combination with the information that is provided.

Thinking in Tromp's [437] terms, a frictional element might fall in the realm of persuasion if accompanied by the proper information. Although some "dark patterns" taxonomies include frictional elements, user-centred design approaches in usable privacy and security have started to use friction to foster reflection in users' interactions [101, 118]. Bringing Tromp's model into manipulative design analysis, the designers in our study do not use "strong" strategies that force users, but they devise other potential paths. However, the "salience" of the mechanism of influence is more subtle. When choice architectures are imbalanced, providing more options to give away your personal data, or if the design strategies appeal to emotions that subvert users' vulnerabilities, the strategy of influence is not salient, nor transparent for the user, and, therefore, manipulative. Exploring these dimensions in their designs becomes crucial for designers to determine when designs are manipulative. A way to explore these dimensions and fix the tensions between principles and UX strategies is provided via design recommendations (See Section 5.6).

Our results, *transparency* and *user experience first*, resonate with the values that Chivukula et al. [83] reported in their study. How they implement those principles ethically seems unclear, justifying unethical implementations on behalf of the greater good. If the impact is positive, designers might interpret that it is justified to create manipulative designs. This is an example of how digital architecture plays a crucial role [275] in "designing with intent". Our results exemplify the borderline cases that Di Geronimo et al. pointed out [111] in their analysis of 'dark patterns'. Based on empirical data collected predominantly from experienced designers, our analysis confirms and expands the findings that Chivukula and Gray [83] started to explore based on data from design students. Our results also are in accordance with previous

literature in design ethics [83, 84, 82], but applied to the use case of influencing behaviours.

5.5.3 From “Dark Patterns” to “Manipulative Designs”

The amalgam of definitions and terms used by scholars to describe “dark patterns” still leaves room for uncertainty, not only for designers but also for policymakers. Several regulations are trying to address the term and the problem they represent, overlooking the tensions that UX designers experience [41]. The coined term “dark patterns” yields problems in two directions. First, it is necessary to listen to the discussion about the reproduction of racist stereotypes when this term is associated with malicious designs [410, 406]. Second, given the term’s vagueness, it sustains confusion among practitioners [410], as reflected in our focus groups. Policymakers should consider its meaning with caution. It is an urgent problem to determine the extent to which “manipulative designs” can be illicit and under which circumstances. The conditions of vulnerability under which manipulation works might be a rationale for regulating these design patterns, including a reformulation of the term. This would allow to provide the term with a precise meaning and avoid confusing the community. As a starting point, we suggest shifting the discussion to a more accurate term, “manipulative designs”, for further reflection in the community: this moves the focus away from specific patterns as the problem concerns an entire system, design strategies and a context of manipulation.

We aim to contribute to the conversation about manipulation in design, restating the problem of dark patterns online, which transcends the ill-intentions of designers and businesses. We have explored how traditional persuasive UX design strategies can easily cross the line of manipulation, turning into manipulative designs. Aiming to understand what design strategies may become manipulative and under which circumstances, we have explored contextual factors that affect designers and may act as a potential trigger of manipulation. This study has helped to consolidate the literature on ethics in design, applied to the specific context of designing for manipulation and showing how the design of manipulative interfaces also depends on context: the personal context of designers that apply ethics, and the context in which the design practice takes place. Therefore, we argue for a holistic analysis when thinking about manipulative designs. First, judging when the strategy is manipulating instead of persuading, and what is the implementation of values through the design strategies. Second, the context in which designers operate and devise their design strategies gains in relevance. Hence, by providing the proper tools, knowledge and management systems within companies, we can empower designers to face the tensions between UX and manipulative designs.

5.6 Design recommendations for practitioners

To facilitate persuasive design, with “apparent” and “weak” design strategies [437], and under consideration of the above-discussed tensions between UX strategies and manipulation, this section provides a set of design recommendations. The recommendations are coupled with questions that designers can ask themselves to assess whether their design crosses the line

between influence and manipulation.

- *Support user goals* [331, 376, 142, 30]. Both manipulation and persuasion seek to influence user behaviour. Trying to influence does not make a design manipulative by default. But if the goal of the design differs from the user's goal, you should assess if your goal also serves their interest. Questions you can ask yourself: Are my design goals aligned with the user's short-term goals but misaligned with the user's long-term goals? Or are the design goals misaligned with the user's short-term and long-term goals? Both cases require reflection regarding the user's interest.
- *Keep incentives resistible* [376, 428, 48]. Designs should not present incentives that are irresistible to specific populations. *Incentives* can be used as long as they do not exploit vulnerabilities. Therefore it is necessary to consider all potentially affected users. Are there user groups that are more susceptible to my incentives? Research vulnerabilities in your target user audience and ask yourself: are the incentives difficult to resist for them?
- *Be careful with emotions* [428, 376]. The design should not appeal to emotions beyond the user's reasonable expectations. If the design appeals to emotions (through language, visuals, sound, etc.), ask yourself: is the emotional appeal difficult to resist? Are there user groups that are more susceptible to these emotional arguments?
- *Use friction for good* [437, 428, 48, 101, 239]. Friction does not always imply 'evil' designs; it can also provide helpful information to users or trigger critical reflection. The key is to analyse *if it supports the users and if users can resist its influence*. So, if the design introduces friction in the user task flow, for example, in form of a pop-up, ask yourself: Is the friction serving the user's interest too? Can they overcome or resist it?
- *Provide fair decision spaces* [296, 428, 335, 281]. To ensure users' agency, all available choices should be presented in an equal manner to the users so that they can *be aware that the options exist*. It is hence indispensable to use adequate signifiers. If the task flow includes decision-making interactions, ask yourself: Does my design present all available choices? Are certain choices emphasised to guide the user's decision-making? Are the emphasised choices highlighting a choice against the users' best interests? Are the emphasised choices highlighting information that is not necessary for the user to make a decision? Is necessary information withheld from the user? On the contrary, if the task flow does not include decision-making, the information design architecture should not lead users towards options against their best interests. You may ask: Is my design employing a mechanism of which the user is not aware? Is the hierarchy of my design prioritising elements to influence the user?
- *Be transparent* [428, 304, 39, 437]. The extent to which users have agency will depend mainly on how transparent you are. Hierarchy and information architecture are essential to creating transparent designs. Designs should avoid barriers to information

and present complete and truthful information so that users can make an informed decision. If the design includes any informational element, you may ask yourself: is the information that I provide complete or only partial? Is the information framed in an unbalanced manner? Is the provided information potentially untruthful? Is relevant information hidden through the information architecture?

5.7 Limitations and Future Work

The objective of this study was to provide an empirical exploration with designers of how persuasive UX practices could become unethical, i.e., manipulative. The present study elicits specific UX strategies that can be considered manipulative; it could be expanded in future versions of this work based on a large-scale set of UX/UI designers. Other expansions of this work might include the analysis of cultural differences in the designers as part of their intrinsic values and perceptions. Although it was not the intention of the study, we included two countries, Luxembourg and the Netherlands, in which the participants worked. We noticed differences in the working style, which might be a nuance to explore in the future.

Similarly, seven of the eight groups were colleagues: they worked together, knew each other and shared the same language. We acknowledge that in those cases, the designers might be shaped by the same organisational practices. We initially tried to overcome this limitation by sampling participants from different backgrounds and experiences so a wide range of organisational practices is represented. In the same way, although participants worked in the same company, they did not always work on the same projects. With this limitation in mind, surprisingly, we found that the group with designers who did not work for the same company had more problems engaging in discussions which might be caused by social desirability bias for ethics-related topics. Further extensions of this work could compare how different companies — from start-ups to large technological companies — shape, constrain or extend the ethical values of UX designers in developing technological solutions. Finally, it will be crucial to expand this work by testing these UX strategies on specific populations that might be considered vulnerable, but also to test under which circumstances persuasive elements might be manipulative on general populations.

5.8 Conclusion

In this paper, we have explained the tensions between UX practices and manipulation, helping to restate the problem of online manipulation and “dark patterns”. We have run focus groups with UX/UI designers and asked them to prototype solutions to influence users online. Building on their discussions, we have elicited the principles that designers follow to design ethically. Furthermore, we elicit the main strategies in UX practices that can become manipulative if the principles are not properly applied. Lastly, we have explored how an imbalanced relationship between designers and stakeholders can contribute to fostering manipulative practices. We provide design recommendations to overcome this tension. This study

provided a new lens to assess UX design ethical practices in the context of influencing user behaviours, showing how manipulation not only results from designers' intentions but also from their circumstances.

5.9 Chapter Takeaways

By studying UX/UI designers practices when asked to influence people, I unveil some specific problems when it comes to implementing interfaces that steer users without manipulative designs. Designers are not sure about the concept of manipulation and user agency, which they implement according to their own views and set of values. The context in which design takes place can determine the tools and knowledge that designers have, limiting their resources to prevent manipulative designs. The standardisation of UX practices and methods — A/B testing to reveal vulnerabilities on users decisions, or aggregated analytics — has led practitioners to normalise the exploitation of users' vulnerabilities with common tools. Their struggles for understanding a free manipulative interface call for better guidance, but raises several questions about to what extent HCI researchers can work to broaden knowledge of users' vulnerability in order to rethink UX practices. It also discusses the level of multidisciplinary involvement — practitioners, legal teams, and HCI researchers — that is needed to fight against the implementation of manipulative designs.

Who is vulnerable to Manipulative Design? A transdisciplinary perspective on the multi-dimensional nature of digital vulnerability.

This chapter is based on the following peer-reviewed publication:

Arianna Rossi, Rachele Carli, Marietjie W. Botes, Angelica Fernandez, Anastasia Sergeeva, Lorena Sánchez Chamorro, Who is vulnerable to deceptive design patterns? A transdisciplinary perspective on the multi-dimensional nature of digital vulnerability, *Computer Law & Security Review*, Volume 55, 2024, 106031, ISSN 0267-3649, <https://doi.org/10.1016/j.clsr.2024.106031>.

As a part of the multidisciplinary conversation that takes place in this chapter, I have contributed to its conceptualisation, methodology, writing, and editing. I have not taken part in the sections related strictly to regulatory analysis (Section 6.5 of the Chapter)

Abstract. In the last few years, there have been growing concerns about the far-reaching influence that digital architectures may exert on individuals and societies. A specific type of digital manipulation is often engineered into the interfaces of digital services through the use of so-called dark patterns, that cause manifold harms against which nobody seems to be immune. However, many areas of law rely on a traditional class-based view according to which certain groups are inherently more vulnerable than others, such as children. Although the undue influence exerted by dark patterns on online decisions can befall anybody, empirical studies show that there are actually certain factors that aggravate the vulnerability of some people by making them more likely to incur in certain manipulation risks engineered in digital services and less resilient to the related harms. But digital vulnerability does not overlap with traditionally protected groups and depends on multifaceted factors. This article contributes to the ongoing discussions on these topics by offering (i) a multidisciplinary mapping of the micro, meso, and macro factors of vulnerability to dark patterns; (ii) a subsequent critical reflection on the feasibility of the risk assessment proposed in three selected EU legal frameworks; the General Data Protection Regulation, the Digital Services Act, and the Artificial Intelligence Act; (iii) and multidisciplinary suggestions to increase resilience towards manipulative designs online.

6.1 Introduction

Dark patterns or deceptive design patterns are design strategies in user interfaces that influence users' decisions concerning the use of their financial resources, of their personal data, and of their time through manipulation, coercion or deception [295, 341]. Dark patterns are a distortion of the UX and UI design strategies that are intended to make the interactions between humans and technologies easier, smoother and more pleasurable.

Even when they are not employed with the intentional purpose of manipulation [169, 379, 86], dark patterns have harmful effects on users, such as loss of autonomy, privacy invasion, financial losses, cognitive burdens, and discrimination, among the others [341, 295]. Even though anybody can fall prey to dark patterns because they exploit cognitive biases [455, 295], there are factors that make certain individuals, groups, or communities more vulnerable than others [340]. In fact, vulnerabilities are transversal in human-machine interactions and are compounded by a set of external and internal factors. Following a general trend in the consumer protection domain that aims at providing a more realistic definition of vulnerable consumers in digital settings [340, 194, 126], in the last few years rising concerns have been voiced about the situations where certain people might be more vulnerable than others [340]. There is a pressing need to identify and denounce the specific, complex, and intermingled conditions of vulnerability towards manipulative design in online services, going beyond the traditional class-based legal views that assign a vulnerability status to specific groups, such as children (see e.g., the European Data Protection Board's guidelines on deceptive designs [41] and that ignores the factors that may render everyone dispositionally vulnerable [194], also due to the way digital architectures are designed [340].

In this article, we intend to identify the sources of vulnerability to dark patterns so that appropriate protections and countermeasures can be devised and effectively applied. We will argue, in line with existing accounts (e.g., [132]), that vulnerability towards dark patterns encompasses two dimensions: (i) the susceptibility to being influenced by the design elements of user interfaces (UIs), and (ii) the severity of the harms and consequences that people may suffer. Such a conceptualization closely follows the notion of risk, which is defined by both the likelihood that a certain event occurs and the severity of the impact it can provoke [179].

In our approach, we seek to show that it is necessary to include a multifaceted notion of vulnerability in risk appraisal methods to enhance their accuracy and efficacy, in data protection as well as in other domains of the law. For instance, both the EU Artificial Intelligence Act (AI Act) and the Digital Services Act (DSA) emphasize the necessity of a proactive technology risk assessment that seeks to develop safe-by-design products and services, while enhancing the accountability of those who conceive, create, and ultimately produce them. Hence, with our analysis of the concrete sources of vulnerability, we intend to make an innovative contribution to the definition of the elements that a sound risk assessment methodology needs to entail.

In order to do so, we will first (Section 2) develop the critiques to the particularistic approach (that identifies certain groups as inherently weaker due to their inner characteristics,

such as age) and universalistic approach (that generalizes the risk of being vulnerable to anyone and thereby carries the danger of levelling out relevant differences that would deserve special protection) of vulnerability [290] to dark patterns and explain why both are partial views that cannot appropriately account for this complex phenomenon. We will then (Section 3) illustrate how the consumer protection domain has pioneered the revisiting of the concept of vulnerability in digital settings.

We will propose in Section 4 an overview of the harms that dark patterns can pose, while, in Section 5, we will explore how data protection law (i.e., the General Data Protection Regulation (GDPR)) and the emerging normative framework designed by the AI Act and the Digital Services Act (DSA), address online manipulative design, vulnerability and the assessment of risks posed by technologies. We will propose in Section 6 a multidisciplinary mapping of the micro, meso, and macro factors that can influence the disposition to harm that people can suffer. Through the discussion of practical examples and empirical results, this conceptualization will show that certain factors may be equal for all while others may disproportionately affect certain people. However, as we will conclude in Section 7, this composite reality adds complexity to the implementation of risk assessment procedures in practice. Conscious of this challenge, this work does not aim to provide definitive conclusions, nor ready-made solutions: rather, it aims to provide the foundational elements that are necessary to evaluate the risks of digital technologies in a more accurate, complex manner.

The contributions of this article are the following:

1. a multidisciplinary mapping of the micro, meso, and macro factors of vulnerability to dark patterns;
2. a critical reflection on the feasibility of the assessment of risk factors of vulnerability to dark patterns, as proposed in the three EU regulations chosen for the analysis (i.e., GDPR, DSA, AI Act);
3. multidisciplinary suggestions to increase resilience towards digital manipulative designs.

The authors have expertise in different domains spanning AI, consumer and data protection law, human-computer interaction, and usable privacy. Their backgrounds range from psychology to social sciences and law. There is also an intersection of the authors' research areas: online manipulation, ethics, and regulation of technologies. The ideas developed in this article result from a fruitful transdisciplinary dialogue across the different domains. The authors position themselves as firm defenders of online users and their everyday interactions with technologies and understand the complexity of their problems as issues that can only be solved with the engagement of all the stakeholders that shape and enable contemporary digital markets.

6.2 Current Conceptualisations of Vulnerability to Manipulative Designs

6.2.1 A Universalistic View: Vulnerability is Inherent to Human Psychology

In the current academic scholarship on dark patterns, there seems to be convergence on the idea that every user of technology can be vulnerable to online manipulation because of common cognitive fallacies shared by all human beings. According to certain accounts [47], dark patterns affect our so-called System 1, which is the complex of mental processes that corresponds to our quick, intuitive, automatic mode of thinking and that enables us to efficiently navigate the numerous, complex decisions we are faced with in every moment of life (see dual process theory by Kahneman [223]). Specifically, the user interface (UI) of a technology can be designed in a way that exploits the vulnerabilities arising from cognitive biases and bounded rationality. Cognitive biases are systematic deviations from the decisions and behaviours that a rational decision-maker would enact [3, 296, 295, 455]; while bounded rationality refers to the limited mental resources to which we can resort to consider all possible courses of action and their consequences when faced with a decision [3].

Both are inherent to human nature and prevent people from acting as perfectly rational agents. As a result, people may overlook, and even consciously ignore, overwhelming information; or select an unfavorable option when presented with many choices, since the cognitive energies and time that can be devoted to each task are limited. For example, studies [442, 162] have demonstrated that privacy invasive design patterns (such as defaults) on cookie banners can play on the status quo bias and significantly increase consent rates. Despite their potential negative impact, cognitive biases also have a functional reason to exist, since they can, together with rules of thumb (i.e., heuristics), facilitate meaningful choices that allow human beings to “satisfice” (satisfy + suffice) [407]. In other words, they enable users to take decisions in a relatively quick manner, instead of getting paralyzed by the pondered analysis of the overwhelming quantity of information, options, and their consequences that they need to navigate in all aspects of digital life.

Identifying cognitive biases as the primary psychological correlate of vulnerability to online manipulation suggests that vulnerability can be minimized or eradicated by overcoming such biases [327, 202], for example by revealing the influence of design elements through increased transparency, or by educating users to recognize manipulative attempts. Such interventions aim at nudging users to engage in more reflective thinking, therefore activating our so-called analytical System 2, as opposed to System 1. This assumption underlies regulatory interventions that mandate information disclosures with the goal of lowering the informational asymmetry between organizations and individuals.

The same assumption also motivates the proposal for cognitive boosts [199, 244] aimed at strengthening people’s competence to make their own choices and for friction designs [316, 101] aimed at making certain actions less easy (i.e., automatic) to perform. Although these

are useful interventions in specific settings for specific goals, they cannot be reasonably embedded in the sheer number of decisions that individuals need to take every day on digital public and private services: after all, our less reflective cognitive system carries the benefit of ensuring that we act quickly and with relatively low effort in our everyday life (Kahneman, 2011). Further, such rationality-enhancing interventions are likely unable to counter hidden deceptive attempts that act below our awareness [45, 111], such as when cookies are installed on a device irrespective of the option selected by the user on the interface (see e.g., Matte et al. [298]). Moreover, they cannot target factors of vulnerability such as e.g., socio-economic differences, as we will detail in Section 6. In conclusion, referring to cognitive biases and the UI design that exploits such biases as the sole elements that can explain the effect of dark patterns on human beings provides a limited view of what is a wider problem space, that hence needs varied solutions.

6.2.2 A Particularistic View: Vulnerability Depends on Inherent Conditions of Certain Groups

A different conceptualization of vulnerability to dark patterns is proposed in the 2023's EDPB guidelines [41] that provide instructions on the design patterns that are not compliant with the GDPR. The EDPB's guidance document recognizes only certain groups as especially vulnerable to online deceptive designs, thereby reflecting a 'particularistic' (or 'class-based' [340]) view of vulnerability. Following the conceptualization contained in the GDPR's Recital 38, the EDPB [41] recalls that dark patterns may have a particularly severe impact on children, since "they may be less aware of the risks and consequences concerned [sic] their rights to the processing" (p. 10). The guidelines also mention "the elderly, persons who are visually impaired, or not as digitally literate as others" (p. 10), who would be less capable of recognizing deceptive designs and less aware of their susceptibility to being influenced.

The EDPB only mentions vulnerable groups again once in para. 44 (p. 20), in relation to the manipulative design practice called "emotional steering" which can allegedly have a bigger impact on those groups that have a "vulnerable nature as data subjects" who may be lured into excessively disclosing their personal data "due to a lack of understanding". Following the GDPR's provisions which mandate that information disclosures should be understandable to children (Article 12(1)) and official guidelines [179], the EDPB's recommendation against such a dark pattern consists in targeted language, tone, and style to raise the chances of understanding and thereby obviate to the vulnerable people's lack of awareness of the risks and consequences of data processing.

We have two main critiques to the EDPB's conceptualization of vulnerability. First, they frame the problem of dark patterns as merely one of informational asymmetry, lack of comprehension and impediment to rational decision-making, which reflects an imperfect understanding of the reasons why dark patterns work. This, in turn, impacts the solutions that are proposed to meaningfully counter their influence, which mainly aim at increasing transparency of data processing operations. Indeed, transparency requirements risk not only to

be shortsighted, but also to have the paradoxical effect of increasing the cognitive burden on users, therefore causing individuals to read and understand the provided information even less (an effect called “the transparency paradox” [329]), let alone use it for more conscious decisions. Second, the EDPB’s formulation of vulnerability only refers to specific groups of users who are intrinsically weak, but it ignores the multifaceted factors that can contribute to a state of vulnerability outside those groups, as we will detail in Section 6. In this context, we do not maintain that children and other vulnerable groups do not deserve special protection, we rather claim that the particularistic approach is an insufficient approach.

6.3 Reframing Vulnerability in Digital Markets

6.3.1 The Reasons why Existing Approaches are Insufficient to Account for Dark Patterns

Taking a universalistic position that ascribes the influence of dark patterns only to common cognitive biases and bounded rationality would minimize the specific susceptibility of certain individuals or groups to dark patterns and the disproportionate effects they may suffer. On the other hand, we are cautious of embracing a particularistic view that identifies certain groups as inherently weaker as if vulnerability was part of their immutable essence, but ignores the multidimensional, dispositional state of vulnerability that anyone may suffer from under various circumstances.

Both approaches provide a non-realistic view of the dynamics to which users of digital technologies are subjected and reflect a partial understanding of the complexity of factors that play a role in such dynamics. Furthermore, the digital environment exacerbates our vulnerability and lowers our resilience. In online settings we tend to act faster than in analog ones, skim rather than read, suffer from shorter attention spans, and more easily trust strangers’ recommendations [343]. Moreover, we are less able to process information, while we resort to rules of thumb and underestimate manipulation more often than in offline contexts [340]. This is why all individuals can be effectively manipulated in digital interactions: vulnerability is common to all human beings because it is rooted in our bodily, limited, perfectible being [137]. Having said that, it is paramount to acknowledge the conditions under which certain individuals or groups may be more vulnerable than others. Reconciling both views is necessary to reliably assess and mitigate the risks carried by technologies and consequently design counter measures that have the desired impact, including the application of effective policy instruments and the appropriate attribution of responsibilities to the stakeholders of digital markets.

6.3.2 Reframing Consumer Vulnerability in Digital Markets

In the consumer protection domain, there have been many recent efforts aimed at surpassing categorical notions of vulnerability and at providing a more complex understanding of

the conditions under which consumers may be vulnerable. This must be interpreted as an attempt to go beyond the static view of vulnerable consumers conveyed by the Unfair Commercial Practices Directive (UCPD) that describes them as “people particularly vulnerable to the practice or the underlying product because of their mental or physical infirmity, age and credulity” (Article 5(3)) as opposed to the “average” or “reasonable” consumer. For example, Baker et al. [17] reject the limitation of immutable internal traits to propose a multidimensional, situational model of consumer vulnerability that is brought about by the interaction of personal states (e.g. mood), personal characteristics (e.g. socioeconomic status), and external conditions (e.g. discrimination) and aggravated by the lack of control and experience that consumers may have.

In digital environments, however, vulnerability can be exacerbated to a novel, worrisome level. In their comprehensive essay, Helberger et al. [194] propose that “vulnerable consumers are not the exception, they are the rule” (p. 180). In the digital society, the authors convincingly argue, vulnerability is (i) architectural, as the digital choice architecture we engage with on a daily basis are designed to infer the vulnerabilities of individuals or even to produce them (e.g., consider the practice of hypernudging [467] which becomes very concerning when it becomes personalized manipulation); (ii) relational, as individuals’ ties to others increase their vulnerability to influencing factors; and is exacerbated by (iii) a general lack of privacy and the concentration of personal data in the hands of few actors. Hence, there are internal vulnerability drivers as well as external drivers in the digital economy, including choice architecture, that make everyone “(dispositionally) vulnerable under the right conditions” [194](p. 194).

Such a nuanced discussion echoes the findings of the European Commission’s (EC) report published in 2016 [126], where the vulnerable consumer was defined as “a consumer who, as a result of socio-demographic characteristics, behavioral characteristics, personal situation and market environment is at higher risks of experiencing negative outcomes in the market; has limited ability to maximize their well-being; has difficulty in obtaining or assimilating information; is less able to buy, choose or access suitable products; or is more susceptible to certain marketing practices” (p. 5) [126]. The EC study highlights that certain factors have a more severe impact on vulnerable consumers than others, such as market related drivers (like the inability to read contract terms and conditions because of small print), behavioral drivers (like impulsivity and risk aversion), as well as situational drivers (like finding it difficult to making ends meet) [126]. Building on this, in 2021 the Commission returned to the interpretation of the concept of vulnerable consumer, defining this condition as situational and dynamic [93]. Thus, it was intended to allude to vulnerability as a modifiable condition, which is not the exclusive prerogative of some, an interpretation that would inevitably leave everyone else immune.

This reconceptualisation of consumer vulnerability in a multi-dimensional key is so well established in the consumer protection domain that it has even been codified in the recent international standard ISO 22458:2022 “Consumer vulnerability — Requirements and guidelines for the design and delivery of inclusive service” for Standardization [144] where consumer vulnerability is a “state in which an individual can be placed at risk of harm during their

interaction with a service provider due to the presence of personal, situational and market environment factors” and that it can be “permanent, temporary or sporadic, long or short term”. Similarly, ISO 31700-1:2023 on “Consumer protection — Privacy by design for consumer goods and services” [145] highlights the influence on the vulnerable state of consumers of “market environment factors” that include “demographic factors, ecological factors, economic factors, socio-cultural factors, political and legal factors, international environments, and technological factors” (p. 6).

In 2023, the OECD has dedicated an entire report to consumer vulnerability in the digital age [340], where it emphasizes that “consumer vulnerability online is increasingly systemic [...] even if at times some consumer groups will continue to warrant specific attention” (p. 6). In addition, it remarks that it may be hard to reach general conclusions about certain groups of consumers, such as the elderly, as vulnerability can be context-specific and influenced by a series of factors, while individuals exhibit varying types of behavior. In conclusion, it is increasingly recognized that a state based view of vulnerability is insufficient to account for digital interactions, which is why this notion is evolving in consumer law regimes around the world [340]. In the EU, this novel conceptualisation could also provide an update to the UCPD’s notion of vulnerable consumers within the framework of the so-called “fitness check” on digital fairness, carried out by the EC [92]. As we will detail in Section 6.6, these factors are relevant for understanding vulnerability to dark patterns as well. For instance, people with mental health conditions may be more susceptible to dark patterns in gambling websites because of their tendency to impulsive decisions and their lower self-control.

6.4 Risk of Vulnerability towards Manipulative Designs and the Harms they Engender

The idea that vulnerability is not a monolithic concept but a rather complex, stratified one is a useful construct to interpret the various imbalances in human-technology interactions afforded by interface design elements, in particular to correctly identify the risks people are exposed to and the subsequent harms they may incur into. This is necessary to determine and implement measures that avoid or lower such risks and thereby contribute to more responsible, vulnerability aware design.

6.4.1 Harms Caused by Deceptive Design Patterns

Various authors have proposed the categorization of the harms engendered by dark patterns [295, 181, 341], but they have not linked them to the sources of vulnerabilities, which is our intent. The OECD [341] arranges the harms in three broad categories, namely (i) those that affect consumer autonomy, (ii) those that cause personal consumer detriment and (iii) those that engender structural consumer detriment. The subversion of individual’s autonomy and decision-making through a more or less overt influence (e.g. a transparent forced action versus a hidden “sneaking” feature) is a constituent element of dark patterns [341, 181]. Individual

harms may include financial losses caused by purchasing unneeded or unsuitable products, receiving items and services of unacceptable low or poor value or quality, spending more than intended, searching for less alternatives [132], and can be provoked by deceptive designs such as subscription traps, unfavorable pre-selections, urgency inducing elements, and confirmshaming. Further individual harms include psychological detriment that depends on emotional distress, such as anxiety [92], cognitive burden due to the spending of unnecessary time and attention [295] (for example, when obstructive dark patterns add unnecessary friction to online processes), including behavioral addictions experienced on social media platforms and videogames.

Then, there are privacy harms, that have been convincingly related to those deceptive designs that are present in consent interactions, exit requests, and user settings [181], even though such harms may be more difficult to identify, quantify, and prove [342], and therefore to compensate, especially if they cannot be robustly linked to concrete financial or material harms or damages. Moreover, data gathering processes may be invisible to users, while their correlated harms may only occur in the distant future [342]. Privacy harms can also cause secondary harms ranging from reputational harms, psychological harms (e.g. embarrassment, anxiety, fear), autonomy harms (e.g. lack of control) and discrimination harms [90]. What makes it worse, is that privacy harms may be impossible to avoid even for informed and careful consumers, as power asymmetries and lock-in effects do not enable users or make it very costly for them to switch between providers [342].

The third category proposed by the OECD concerns the harms that have “a cumulative impact on consumers collectively, even where they [are] imperceptible harms at the individual level” (p.26) [341], namely weaker and distorted competition and loss of consumer trust. In this regard it must be noted that often the harms that consumers endure online are mostly embedded in micro-transactions which, considered individually, may per se not be sufficient enough to motivate a consumer to take action, denounce the malicious deed and report it to the relevant authority or ask for redress, regardless of the fact that the cumulative effect may be quite big or extensive. Moreover, there often needs to be a minimal threshold of materiality, significance, or severity of damage present before one can obtain redress, as is for example, the case with GDPR infringements [181]. Hence, even in the rare cases when users are aware that they have been impacted, it may be too expensive to seek justice for individual dark patterns they encounter in online interactions.

6.4.2 The Two-fold Vulnerability to Dark Patterns’ Harms: Likelihood and Severity

There are at least two ways in which vulnerability threats can be experienced: the susceptibility to the influence of dark patterns and the severity of their effect [132]. First, although everyone is dispositionally vulnerable to the influence of manipulative design patterns because of common cognitive biases or market conditions, some people under certain circumstances are more likely to be influenced or deceived. For example, evidence shows that the risk is higher

when they have lower educational levels, are older, or are under time pressure [92].

Second, although everyone can be harmed by the use of dark patterns, for someone the detriment may be more severe, for instance because they are less able to recover from a negative experience, like spending more than they can afford, or because they may have less access to remedies [132]. For instance, recent statistics published by the French data protection authority show that most complainants in 2021 were managers and had an elevated level of education, which may be correlated to the digital skills that are needed to submit complaints online [100]. This shows that there are variations in the extent to which people can uphold their rights. Moreover, both primary and secondary harms can exert varying impacts on different people. For example, survivors of domestic abuse may also suffer from physical harm because of the reckless use of privacy-invasive defaults in certain applications that broadly share one's people position and other personal details by default, often in an invisible manner to the user [358]. The two-fold nature of vulnerability closely recalls the definition of risk which is determined by its likelihood of occurrence and the impact it may have [227].

6.5 Risk Assessment and the Conceptualization of Vulnerability in the Regulation of Manipulative Designs

6.5.1 Risk Assessment as Safeguard in Technology Development

Being able to determine how harms may disproportionately affect certain individuals or groups is an integral part of the process of evaluation of risk which encompasses the identification of the risks, the evaluation of their impact and the establishment of appropriate mitigation measures. Risk assessment is indeed a central instrument in various European legislative instruments that regulate the development and deployment of technologies, such as those that process personal data, digital services and artificial intelligence (AI) [127].

Assessing when the design of technology becomes problematic is a necessary preventive approach that can address at least two challenges raised by dark patterns. Firstly, designers and developers do not employ manipulative designs only when they want to intentionally deceive users, but also when they have good intentions [379, 83]. Secondly, the distinction between illegitimate manipulative patterns and legitimate persuasive design techniques that are commonly adopted to steer individuals' actions towards intended goals is so subtle [169, 379] that it raises ethical questions. Designers can often not discern whether what they are doing is persuasive or manipulative [379, 83], for instance in the case of emotional design [333] and friction design [101]. The lack of clearcut distinctions makes it hard to regulate interface design with straightforward rules and blacklists. What constitutes a manipulative, illegitimate (and even unlawful) design element may be context-dependent and, in some cases, difficult to gauge objectively. This is where a risk-based regulatory approach results more useful than a rule-based one, because it can better adapt to the ever-evolving nature of digital technologies and is therefore more flexible and future proof [62].

In the following sections, we briefly analyze three EU regulations that (i) introduce risk

assessment to identify the mitigation measures against the risks entailed by emerging technologies, services or processing operations; (ii) define vulnerability; (iii) constitute relevant legal instruments to contrast dark patterns, as they contain provisions against manipulative digital interface designs. The selection is premised on the General Data Protection Regulation, the Digital Services Act and the AI Act. The analysis intends to show that, first, digital manipulation through the design of technologies is a growing regulatory concern for the EU policymakers who make attempts to prohibit problematic design practices in favor of fairer ones; second, the three legal instruments reflect that the notion of vulnerability is evolving from a rigid particularistic view into a more nuanced approach, but without necessarily converging; third, it is necessary to compose a realistic, layered perspective on the concept of digital vulnerability to be able to assess risks in an accurate manner.

6.5.2 The GDPR

Vulnerable groups in data protection law

Data protection law is meant to shield people from data practices that may erode their rights related to their personal information and their freedoms or weaken their decision-making ability [68], which is what privacy-related dark patterns indeed seek to achieve. As anticipated in Section 2, the GDPR explicitly acknowledges that certain groups deserve strengthened protection, namely children, because of their lower awareness of the risks inherent to data processing and of their rights (Recital 38). Such a class-based perspective is reinforced by official interpretations of the legislation, like in the EDPB's Guidelines on Data Protection by Default and by Design [42] and the already cited Guidelines on Deceptive Design Patterns [41], where there is an emphasis on the necessity to provide "specific protection" to children under 18 and other vulnerable groups. This perspective recalls consumer protection views where children are regarded as vulnerable consumers because of their lack of experience and their lower ability to resist influence [340]).

Contextual risk assessment to account for nuances of vulnerability

The lack of awareness and understanding of the consequences of data processing and the existence of legal rights can befall anybody. In fact, as noted by Malgieri and Niklas [290], the GDPR also includes a more nuanced view of vulnerability, since the notion of risk assessment is central to enable the effective protection of the people whose data is processed. If we understand vulnerable people as those exposed to higher risks of damages, then the risk-based approach in the GDPR "can play a significant role in recognizing and conceptualizing the variety of risks (and layers) that can amplify, expose and exploit different vulnerabilities" [290] (p.11). Indeed, the risk analysis carried out by the data controllers needs to consider the "varying likelihood and severity [of the risks] for the rights and freedoms of natural persons" (Article 24), namely how the risks related to data processing may concretely exert different impacts under different circumstances. With the goal of yielding the appropriate measures

for the mitigation of the identified risks, such an assessment must be contextual by factoring in the specific nature, scope, context, and purpose of the processing, and continuous by considering the state of the art [42].

Further, the processing of personal data of vulnerable individuals may result in a high-risk activity for the involved individuals. This is one of the conditions under which a Data Protection Impact Assessment (DPIA) must be carried out (Article 35) to enable the design of suitable, contextual mitigation measures. The guidelines on DPIA (Article 29 Data Protection Working Party, 2017) contain a non-exhaustive list of specific vulnerable people such as children, employees, and “vulnerable segments of the population” such as mentally ill persons, asylum seekers, the elderly, and patients. Nonetheless, they also include power imbalance as a factor of vulnerability, which may make certain people ‘unable to easily consent to, or oppose, the processing of their data, or exercise their rights’ (p. 10) [179], for example in the employee-employer relation. In this regard the GDPR is clear, as for example consent for the processing of data where there is a clear imbalance between consentees and the organizations requesting consent should be examined very closely, as consent should be freely given (Recital 43 GDPR).

Power asymmetries in digital markets are widespread, though, and can not only be redressed through enhanced transparency about the processing operations and purposes, since the latter seeks to address informational asymmetry. Rather, manipulation can be countered by fairness, the main principle of data protection violated by any deceptive design [41]. Fair data management excludes any processing that is “unjustifiably detrimental, unlawfully discriminatory, unexpected, or misleading to the data subjects” [42](p. 18-19), and avoids the exploitation of their needs and vulnerabilities [42], the latter being a constitutive element of manipulative practices [429].

The DPIA is likely to be mandatory also in the case of automated decision-making with legal or similar significant effects (Article 22). Dedicated official guidelines [355] have expanded the notion of vulnerable groups beyond children recognizing that “[p]rocessing that might have little impact on individuals generally may in fact have a significant effect for certain groups of society, such as minority groups or vulnerable adults” (p. 22). When the effects of automated decisions on individuals are assessed, an important factor will be whether the data controller used “knowledge of the vulnerabilities of the data subjects in a targeted way, such as people in financial difficulties targeted with adverts for high-interest loans.

Legal provisions against dark patterns

The GDPR does not contain explicit references to online manipulation, although it has introduced behaviorally informed provisions aimed at contrasting the use of deceptive design techniques that affect individuals’ data privacy. For instance, the principle of data protection by design and by default enshrined in Article 25 seeks to minimize data collection, storage and use as the default situation. Together with the obligations related to data minimization (Article 5(1)(c)) and purpose limitation (Article 5(1)(b)), Article 25 attempts to shield individuals from function creep and abusive data-hungry practices by countering the status quo bias, accord-

ing to which people tend to stick with the default option provided to them (i.e., the path of least resistance) (see Section 2.1).

Similarly, the notion of unambiguous consent (Article 4) is paramount to combat default effects. In 2019, the Court of Justice of the European Union in a landmark case has provided the interpretation that pre-ticked boxes cannot signify a legally valid consent, because the user is not actively engaged in the decision. Further, the obligations concerning transparency (Article 12), which also pinpoints the notion of informed consent, should also be interpreted as a tentative to contrast hidden, obscure or misleading data practices. Moreover, as mentioned in Section 2, the GDPR mandates to adapt the language and style of communications to children and, more broadly, to any intended audience [179]. In general, though, it is fairness the foundational principle that is violated by any dark pattern [41] since fairness is meant to ensure that personal information is “not processed in a way that is unjustifiably detrimental, unlawfully discriminatory, unexpected or misleading to the data subject” (p. 12) [42].

A combination of universalistic and particularistic views

In conclusion, in the GDPR, the concept of vulnerability is both tied to a ‘particularistic’ approach that explicitly mentions children and other groups, as in need of strengthened protections, and to a ‘universalistic’ approach that is based on contextual risk management. This is why Malgieri and Niklas [290] fruitfully adopt Luna’s idea of ‘layered vulnerabilities’ [283] and transpose it to the data protection domain, where everyone is deemed vulnerable because of the general “inferiority, dependency, and subjugation of individuals in the context of processing data” [290](p. 16), but where some are more vulnerable than others, depending on internal and external factors.

6.5.3 The Digital Services Act

Definition and prohibition of dark patterns

The Digital Services Act (DSA) sets the groundbreaking record of being one of the first EU regulations that explicitly refers to ‘dark patterns on online interfaces of online platforms’, defined in Recital 67 as ‘practices that materially distort or impair, either on purpose or in effect, the ability of recipients of the service to make autonomous and informed choices or decisions’.

The DSA aims to protect individuals’ autonomy from the undue influence of online intermediaries, which is one of the main harms engendered by dark patterns (see Section 4), by prohibiting deception and manipulation in the design, organization, or operation of the online interfaces of platforms (Article 25 (1)). However, the scope of application of these provisions is still debated, as Article 25(2) clearly excludes the practices already covered by the UCPD and the GDPR, wherein most dark patterns fall under the scope of those two regulations [267]. Article 25(3) grants the power to the Commission to develop guidelines clarifying how Article 25 (1) applies to the asymmetric presentation of choices, that is to say assigning (visual, auditory, etc.) prominence of certain options over others and making it more difficult and time-

consuming to select certain options; to nagging users with repetitive requests which interfere with the user experience; or to making the cancellation of a subscription overly difficult.

Systemic risk assessment of manipulative designs

Although the DSA does not explicitly acknowledge that dark patterns may have layered effects on different users, it does recognize the necessity of contextual risk assessment when the design of widely adopted digital services can impact the vulnerability of people. Of interest is the introduction in Article 34 DSA of the obligation for very large online platforms and for very large online search engines to assess systemic risks arising from content moderation, recommender systems, advertising, and other parts of the design of their services. This duty is combined with the obligation of applying reasonable, proportionate, and effective mitigation measures (Article 35) that include the adaptation of the design, features and functioning of the services, including their online interfaces (Article 34(1)a).

6

Vulnerability to online manipulation

One of the categories of systemic risks identified in Article 34 DSA concerns the impact of a service on fundamental rights, including the right to privacy and data protection, non-discrimination and consumer protection, and the rights of children, among others. The emphasis is placed on the specific risks that children may incur into, which may depend on the “design of online interfaces which intentionally or unintentionally exploit weaknesses and inexperience of minors or which may cause addictive behavior” (Recital 81) and impairs their health, physical, mental and moral development. This does not merely refer to the impact of dark patterns on children but to that of online manipulation in general, which includes the harmful effects of online hate speech (Recital 62), advertisement (Rec. 95), and disinformation (Rec. 104). In line with such a reading, Article 28 introduces an obligation for providers of online platforms to “ensure a high level of privacy, safety, and security of minors”.

Although minors may be particularly exposed to the nefarious consequences of addictive and exploitative designs, all adults may be susceptible to them, given that they may be similarly inexperienced or unable to protect themselves from the effects of online manipulation. The DSA has an opening in this respect: Recital 83 acknowledges that an additional systemic risk (identified in Article 34(1)(d)) derives from the “design, functioning or use, including through manipulation, of very large online platforms and of very large online search engines” that can engender serious negative consequences, for instance on a “person’s physical and mental well-being, or on gender-based violence”. Such risks may also originate from “online interface design that may stimulate behavioural addictions”, such as design patterns meant to make users increase the time they spend on a certain service.

Only in relation to targeted advertisement there is a specific reference to individuals’ vulnerabilities (Recital 69), because it may negatively impact certain groups and amplify societal harms. This is why profiling for marketing purposes cannot be based on sensitive data (Article 26 (3)). Since technological evolution and big data gathering will make it increasingly easier

to tailor dark patterns to target specific user characteristics and vulnerabilities [341, 340, 92], thereby increasing their potential for harm, an anticipatory perspective would caution to foresee and mitigate such developments proactively even beyond advertisement.

Contextual, empirically-based risk assessment

It is difficult to anticipate at this point how the risk assessment should be carried out in practice and how to ensure that it accurately accounts for vulnerabilities. Article 34(2) provides a non-exhaustive list of elements to factor in the risk evaluation. Yet, the focus is on the internal design features of the system, while the other elements that could contribute to the risk (e.g., users' digital literacy) are only vaguely alluded to in Recital 84 [293]. That said, Recital 90 shows an interesting opening to the engagement with impacted stakeholders: it suggests that very large online platforms should conduct their risk assessments, and design their risk mitigation measures "based on the best available information and scientific insights and that they test their assumptions with the groups most impacted by the risks and the measures they take" and "with the involvement of representatives of the recipients of the service, representatives of groups potentially impacted by their services, independent experts and civil society organisations" through "surveys, focus groups, round tables, and other consultation and design methods". Such a participatory assessment must be carried out periodically and "in any event prior to deploying functionalities that are likely to have a critical impact on the risks identified" (Article 34(1)).

An opening towards an empirically-based determination of vulnerabilities for accurate risk assessment

To sum up, the DSA requires very large online platforms and very large search engines to carry out a contextual, systemic risk assessment to avoid all forms of online manipulation, including dark patterns, when an impact on fundamental rights is foreseeable. Albeit the regulation mainly identifies internal risks deriving from the design of the system, it also opens up to the necessity of involving the impacted stakeholders to provide data-informed and scientifically grounded measures for risk appraisal and mitigation. This indicates sensitivity towards the concrete impact that digital services, especially very large ones, may have on various consumers. However, digital services that do not fall under such categories are exempted from such obligations while the scope of application of the prohibitions contained in Article 25 is narrow.

6.5.4 The Artificial Intelligence Act

Towards a more nuanced understanding of vulnerability

Although the Artificial Intelligence Act (AI Act) Proposal launched in April 2021 adopted a class-based interpretation of vulnerability dependent on age or disability status, the most recent version, approved by the European Parliament in March 2024 aims for a more nuanced view of

vulnerability. In line with the GDPR's provisions on automated decision-making, Article 5(1)(b) prohibits AI systems that exploit human vulnerabilities, due to their age or disability, or a specific social or economic situation, to materially distort people's behaviours "in a manner that causes or is reasonably likely to cause [...] significant harm".

Such an approach demonstrates the recognition that even certain situations, and not only a status, can expose people to the manipulative dynamics of AI systems, such as the emotional steering [397] and cuteness [255] that e.g., social robots and conversational agents can leverage to incentivise certain behaviours (e.g., purchases). What is even more worrying is the hyperpersonalized manipulation that AI systems can engender: there are growing concerns that e.g., smart devices can profile their users through the unique interactions that occur between them and then leverage that knowledge to enhance the effectiveness of their influence [94] by targeting specific vulnerabilities. Considering these concerns, the AI Act misses the chance to expand on the notion of vulnerable people. Unlike the broader definition of the AI Act text approved by the European Parliament June 2023 that included personality traits and reference to abilities rather than disabilities, the current definition in Article 5(1)(b) replicates a classification based on specific conditions that may be transitory or permanent, but are nevertheless always proper of certain pre-classified groups, thus not easily generalizable to other cases.

The prohibition of subliminal and manipulative techniques

Article 5(1)(a) uses ambiguous wording to prohibit the use of AI systems that deploy "subliminal techniques beyond a person's consciousness" (i.e., stimuli that people cannot perceive, see Recital 16) or "purposefully manipulative or deceptive techniques" which distort people's behaviour by appreciably impairing their ability of making informed choices. Such expressions are not precisely defined within the document, but the AI Act qualifies such techniques by their effect, since they would direct individuals to take decisions that they would have not taken otherwise and that cause or could reasonably cause significant harm. By adding reference to manipulative strategies that can, but must not, be below the level of consciousness to effectively distort autonomy and by recognising that harm can have a multifaceted nature, the latest amendments reflect a more comprehensive and realistic view of the many ways AI systems can unduly influence human beings.

However, technologies that implement AI are often designed in a way that humans can interpret and understand, such as the humanoid traits assigned to artificial agents to enable smooth, natural interactions (e.g., the unnecessary eyes and human-like voice of social robots) [71]. This is where manipulation can be functional: attributing human physical and mental characteristics to nonhuman entities (i.e., anthropomorphism) helps people to interpret the actions of computer and robots and to create mental models Seymour and Van Kleek [397] that enable meaningful interactions. This and other sorts of "banal deception" are inescapable for the acceptance and integration of AI systems into our everyday lives: it is hence purposefully engineered into devices like social robots and voice assistants [71, 323]. As a consequence, adopting a literal, but plausible, interpretation of Article 5(1)(a) would have the para-

doxical effect of characterizing all AI systems as deceptive, and of reducing the assessment of their prohibition only to the severity of any damage produced (or producible). However, how to determine whether a harm is significant is still open to debate. Conversely, since AI applications generally circumvent human rationality or at least act below the awareness level [35], the harms that are inherent to subliminal techniques appear difficult to identify and quantify, thereby threatening the applicability of Article 5(a).

6.5.5 A Selective and yet Uncertain Risk-Based Approach

To sum up, the AI Act adopts a hard approach to AI technologies that have the purpose or effect of infringing individuals' autonomy and cause significant harm through manipulation or exploitation of their vulnerability. These are prohibited, as they pose an unacceptable risk to the health, safety, and fundamental rights of individuals. Apart from prohibited systems, the AI Act also foresees certain categories of AI systems that are classified as high-risk. Whether an AI system falls under the category of high-risk has been predetermined by the legislator (in Annex I, Annex III, and Article 6) and must be assessed to determine whether the harm they can cause is "significant". This is a challenging aspect, since when the harm cannot be classified as significant, including when it does not materially influence the outcome of decision-making (Article 6(3)), an AI system is not regarded as high-risk and is thus subject to less stringent oversight, since there is only a general obligation to transparency.

This implicitly suggests that if the artificial nature of the generated outputs is made clear, the adverse effects of an AI system are automatically reduced [263]. However, given that dynamics and features can influence individuals in subtle ways, the mere awareness of the artificial nature of the system or of the outcome are not sufficient to prevent possible manipulative drifts [71, 454]. Therefore, while the subdivision into risk classes seems to be capillary, determining the level of harm, specifically in terms of manipulative power, is still critical. The uncertainty on the assessment of such harms [182] casts doubts on the effectiveness of the AI Act [125, 474, 422, 416] and calls for the determination of the actual drivers of vulnerability to digital manipulation.

6.5.6 Brief Conclusions on the Three Regulatory Approaches

The GDPR and the DSA specifically refer to children as a vulnerable category of individuals that require additional protection due to their inexperience, susceptibility to influence, and other internal characteristics regarded as proper of non-adults, which constitutes a particularistic approach. However, both instruments introduce risk assessment as a mandatory measure under certain conditions of potential harm, which suggests a layered approach that takes into account the composite nature of vulnerability in digital environments. The AI Act also attempts to move away from a particularistic approach by prohibiting AI systems that exploit human vulnerabilities that concern personal traits such as age and disability, as well as situational factors that may be temporary.

In all three regulations, the evaluation of risk appears to be based on human characteristics that encompass both internal and external drivers of vulnerability, as well as on the design features of the technological system at hand. Following the tendency observable in consumer protection policies (see Section 6.3), the concept of vulnerability to digital technologies starts moving away from a class-based perspective, even though this tradition is still entrenched in the three normative instruments under analysis. Moreover, the GDPR and the AI Act find it relevant to make a distinction, even if implicit, between vulnerable subjects and those that can recur to rationality to shield themselves from manipulative techniques, for example, thanks to increased transparency on the logic involved or on the output of the AI-based decision. Thus, laying down a reasoned, realistic mapping of the various drivers of vulnerability to manipulation exerted by interface design in digital settings becomes necessary for guiding organizations in their risk assessment practices.

6.6 Factors Influencing the Risk of Vulnerability to Manipulative Designs

6

This section will illustrate the elements that must be factored in the evaluation of the threats posed by deceptive design patterns. We will complement vulnerability theories with the scientific perspectives of human-computer interaction and ecological psychology, that maintain that the immediate surroundings of users influence their perception [195, 160] and their actions [335]. Their environment impacts their use of technologies, which can lead to vulnerability. For instance, a cookie banner that presents the rejection option with an inconspicuous link instead of a salient button might prevent users from noticing it, especially those with visual impairment, leading them to accept privacy-invasive cookies that weaken the protection of their personal data. We posit that understanding how technologies can trigger vulnerabilities can usefully support risk assessment practices and the identification of the ways people can become resilient to technology-mediated manipulation harms.

To understand the different factors that can drive vulnerability to deceptive designs we leverage the Ecological System Model [57] and ecological theory of affordances [195, 160]. The way people perceive, interact, and experience technology depends on their environment in every system, as defined by Bronfenbrenner [57]: micro, meso, macro, exo and chronosystem [58, 349, 220, 195]. The combination of the elements belonging to such systems increases the likelihood of being affected by manipulative designs as well as their impacts. In the domains of human-computer interaction, computer-mediated technology and computer-supported collaborative work, this framework has been expanded to include the mediated aspect of technology through the interactions or informatics layer [320], or the techno-subsystem [220]. This informatics layer mediates how users understand and interact with technology: deceptive designs are situated in this layer, and therefore, mediate the opportunities for vulnerability through the interaction.

We have taken inspiration in how this model has been applied in HCI and the design of

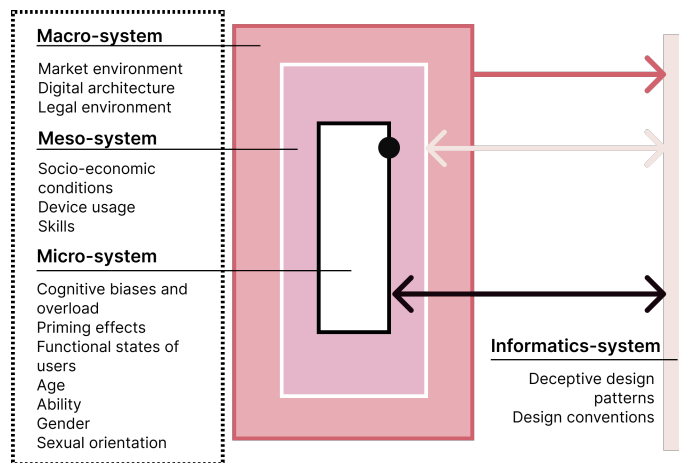


Figure 6.1: The ecological systems theory model applied to the development of vulnerability to deceptive design. As the arrows show, the factors pertaining to the macro-system have an influence on the informatics-system and deceptive design deployment. The factors of the micro-system and the meso-system have an influence on the informatics-system and vice versa. At the same time, factors of the meso-system also influence the micro-system by mediating the users' interaction with deceptive patterns in the informatics system, represented with the dot.

technologies in contexts like design for healthcare [97], personal informatics [320] or socio-digital inequalities [195], and we apply it to the context of interaction with deceptive designs. As shown in Figure 6.1, the macro-system includes the macro conditions where users and technology interact, like the economic and regulatory systems (see Section 6.6.1). Meso conditions pertain to the environment in which the interaction with technology occurs (e.g., communities and neighborhoods), namely the “everyday social, physical, and technical environment in which people live their lives” [195] (p. 24) (see Section 6.6.2). The micro-system refers to the individual and their specific personal conditions of users, such as temporary states like stress or fatigue, or more or less permanent states, such as mental health conditions (see Section 6.6.3). Exosystems are events or structures that indirectly affect the user, like the user's family workplace [220]. Chronosystems refer to how temporal changes in the environment influence the individual.

Since exosystems and chronosystems are related to other conditions pertaining to other systems, such as socio-economic status or age, and they add an unnecessary layer of complexity to our model, we find it useful to confine our discussion to the micro, meso and macro systems which already provide a helpful framework of reference to understand vulnerability to deceptive designs. We situate deceptive designs in the informatics-system that is influenced by the macro-system. Moreover, the informatics-systems have a bidirectional influence over micro and meso-system, since deceptive designs affect the individual interaction of users, but at the same time factors of the meso-system shape this relationship and understanding of deceptive designs.

For us, vulnerability to deceptive designs occurs when one or more factors operate, giving rise to specific contextual and situated experiences that makes users vulnerable, following the ideas of drivers of vulnerability from Malgieri [289]. Hence, our intention is not to make an exhaustive list of conditions that strictly belong to one or another category, but rather to illustrate the complexity of vulnerability as a human condition within the (online and offline) systems where humans live and interact. In the following sections we explain how the interrelation of factors in every system contributes to experiences of vulnerability to deceptive design patterns.

6.6.1 Macro-Factors: Interface design is Never Neutral. The Role of Digital Architecture in Increasing Vulnerability

The digitalization of all aspects of society such as banking, commerce, leisure, healthcare, and public administration is forcing individuals to continuously engage in digital transactions. The asymmetries with digital services and the very nature of the market [194] put individuals in situations of vulnerability [340], even though some people may be in a riskier position than others, given, for example, their level of credulity, or their difficulties navigating the Internet. For instance, in the context of online manipulation, older adults have shown the desire of disengaging from technology, reclaiming a right not to be bothered by manipulative practices. They express frustration because they feel like outsiders in the use of imposed technologies [383]. Market conditions and modern society's tendencies also result in a race to the bottom that causes dark patterns to be omnipresent in digital services. The concentration of power and information; the widespread reliance on personal data collection aimed at offering personalized advertising online; and the pervasive use of e-commerce services and social media are some of the market conditions that ease the proliferation of dark patterns. For instance, a 2022 European Commission report [284] shows that almost all the most popular applications in the EU contain at least one deceptive design. In such settings, the complexity and interconnectedness of digital products embody higher risks [340], that can additionally increase the power asymmetries when users become dependent on such products [194]. For example, because their social networks heavily rely on a certain product [292] or because their employer mandates such use.

The set of actors, norms, architecture, and market that constitute the digital architecture is conceived to shape user interactions and afford certain actions, thereby impacting the behavior of individuals [269, 275]. Thus, the emergence of potential deceptive elements that aggravate the vulnerability of users are determined by the various stakeholders that participate in the design of technologies at large, the position of the designer within the organisations that generate and commercialize certain designs, and the context in which the design activities take place [83, 379, 462, 164, 194, 275, 48, 86, 275].

In addition, the legal norms that regulate digital architectures also influence the protections embedded in the design of digital services, as the example of privacy-preserving defaults illustrated in Section 5.2 shows. But they may also engender the paradoxical effect of af-

fording certain deceptive practices. For instance, when transparency requirements produce off-putting lengthy explanations and long lists of options that fatigue readers (the so-called “transparency paradox” [329]) rather than enhancing their autonomous, informed decision-making capabilities.

6.6.2 Interaction between Meso-Factors and Informatics-System: Users Perceive and Resist Deceptive Designs Differently

People interact with interfaces through “affordances” and “signifiers” that enable, or disallow them to interact in specific ways [335, 98]. “Affordances are opportunities for behaviours” [253] (p.189); therefore, certain interface design elements “afford” users different opportunities. However, these possibilities for action must be highlighted with “signifiers” to be perceived by users, for example, to discover more information, indicate or signify a choice, or accept or refuse the terms of a contract. In this section, we explore (i) how the design of the interface, embedded in the digital architectures, might increase the conditions of vulnerability and exploit them (ii) and how users’ material conditions impact how they perceive and interact with online services.

Design conventions: Visual perception based on salience and spatial organization of the elements on the screen

When it comes to graphical user interfaces, visual salience and cluttering are the main factors that channel attention to certain areas of the visual field [200]. The concept of visual salience describes the features of an object or region in a visual space that is distinct from its surroundings. Cluttering refers to the number of objects in that space and the complexity of the information organization, which makes visual tasks, such as searching for a stimulus (e.g., like a link or button on the screen), more difficult [200].

These two concepts can be engineered to drive users’ attention towards predetermined choices [13]. Studies show the influence of visual salience on users’ shopping task performance [425], on choice [311] and recall [450] and on decision-making processes in general [217]. Cluttering can distort the processing of information and nudge the selection of one option over the other because it raises users’ cognitive load such as amount of mental effort required to process and understand information. As a result, users may divert from activities not aligned with the system designer’s goals. For example, the “disguised ad” patterns play on visual salience [169]. The user can be misguided by the false similarity between the elements of the interface and an external ad, and consequently, unintentionally click on it. Similarly, Bösch et al. [47] identify the wall of text of privacy policies as a dark pattern that overburdens readers and can make them abstain from engaging with the legal terms.

Visual salience of certain interface elements can be used to present options in a way that can nudge people [169] to select certain products or services over others or to take privacy invasive decisions instead of privacy friendly ones. Somehow complementary to this strategy is the low contrast between critical text information and layout, which makes certain options

available but barely visible. For example, consent refusal on cookie banners or unsubscribe options from newsletters, therefore more costly for users in terms of time and cognitive effort. Jarovsky [216] also adds the contrast effect that plays on subtle color and contrast schemes to give or conceal visual saliency to some elements.

Even though the visual elements and how the choice architecture is presented in the user interface matter, there is no deterministic effect between one design element and the influence on behaviour, since personal factors also play a role. Although the literature that demonstrates the effects of manipulative designs is still growing [171], we can leverage some evidence concerning how manipulative design elements affect the users' choices online. Some studies have demonstrated that removing option buttons increases users' cookie consent [336, 162] or privacy-invasive decisions when dark patterns are combined [280]. Luguri and Strahilevitz [280] found that playing with the information shown to the user (i.e., hidden information) and the availability of options (i.e., default choices), doubled the users' acceptance rates on subscriptions, while highlighting the "accept" option had no significant impact. Berens et al. [31] found that, in cookie consent banners, the reject option as a link instead of a button changes the responses significantly. They also found a weaker impact when the "accept" button is highlighted and the "reject" button is next to it. Although there is still room to disentangle the effects, the experiences of users converge: users find it hard to identify manipulative design elements [280, 45, 287, 163] and, even when they are aware of their presence, feel powerless to counter them [163, 45, 46]

Socio-digital inequalities as internal meso-factor

Users' perception also depends on their immediate surroundings and other contextual factors. Belonging to a group that shapes their identity implies a set of values, norms, and attitudes that influence how users interact with technology [195]. When different conditions of class, education, race, or gender intersect [98], they may give rise to an intersectionality of vulnerabilities in the online realm. The outcomes of the use of technology (i.e., the extent to which users can take economic, cultural, or well-being benefits from technology) are determined by socio-economic inequalities [391, 443, 198] that are referred to as "digital inequalities" [198].

The experience and skills that are needed to navigate the digital environment are determined by formal and informal education, including family environment, support from friends, and personal attitudes [446]. These factors are all strictly related to socioeconomic conditions and impact how people process, understand, and relate to online and digital information. People living in an underprivileged state with less material, temporal, social, or cultural resources, find more complicated to develop digital skills to protect themselves online and to cope with harms afterwards [198, 187, 151, 134, 196, 446]. Socio-digital inequalities are a meso-factor that in combination with the rest of the systems drives vulnerability.

As defined in the UCPD (see Section 6.3), "credulity" is one of the conditions of the "vulnerable consumer" and may be related to the level of education and digital literacy. Credulity may prevent users from understanding interfaces, the potential risks and harms from online interactions, and the business models that shape the design of an online interface [46, 340].

On the contrary, living as a highly educated person in a resource-rich environment will enable individuals to be more cognizant of the benefits that the internet provides [198, 445, 134], as well as enable them to know how to be protected from online harm.

Prior studies show that the educational level of people plays a role in the resilience of people against dark patterns. For instance, Bongard-Blanchy et al. [45] found that people with an educational level lower than the Bachelor degree level were less likely to identify dark patterns online, while Luguri and Strahilevitz [280] showed that a lower level of education increased the likelihood of accepting privacy-invasive options. Conversely, Zac et al. [472] did not find significant results that relate education and income to dark patterns resistance.

These studies provide preliminary indications that certain socioeconomic conditions can strengthen the influence of interface design, which in turn can aggravate the state of vulnerability, for example when it weakens people's privacy. Thus, DiPaola and Calo [113] point at the idea of socio-digital vulnerability: the environment mediates creating more opportunities for exploiting vulnerabilities and create new threats. Similarly, when Sánchez Chamorro et al. [382] explain the social component of resistance to manipulative design of teenagers at risk of social exclusion, they highlight the impact of socio-digital inequalities as a reinforcing factor of vulnerability online. Teenagers learn about the deceptive designs' risks from family and friends, who also help them to cope with the effects of these designs. However, if the environment is less educated or acquainted with technology, the effect might be the opposite. Similarly, the Stigler Committee [478] also notes that dark patterns imposing transaction costs, like cumbersome opt-out options, may particularly impact less tech-savvy users such as elderly or less educated people.

The socio-economic status also influences the use of devices since mobile phones are the most common devices among individuals pertaining to lower SES families [438]. Given that manipulative and addictive elements present in mobile applications are varied and pervasive [180], they are more likely to affect individuals from lower socio-economic classes [265]. Similarly, Radesky et al. [367] report how children of families in a lower socio-economic situation would be more likely to find mobile applications with manipulative designs. This may hint at a greater risk for those that mostly or exclusively recur to mobile devices to access digital services, which may predominantly coincide with a certain socio-economic status.

6.6.3 Micro-Factors: Internal Human Factors that Influence Vulnerability Online

There are a number of human cognitive and perceptual elements that can influence the predisposition to being vulnerable to dark patterns. In this section, without the pretense of being exhaustive, we will analyze individuals' features that encompass cognitive biases and cognitive overload, priming effects, functional states and other personal conditions. Some of these factors may be permanent, some may be temporary, and some may evolve over time (e.g., age).

Cognitive biases and cognitive overload

As contemplated in Section 2, most of the literature in this domain focuses only on complex high-level psychological features, namely on cognitive biases [47, 455, 296, 280] and psychological needs [47]. It has been noted that, even though cognitive biases are intrinsic to the human nature and are exacerbated by the online environment, not everyone suffers from all these biases, or we may do so with great variation [132].

Many cognitive biases and heuristics have been identified as having a primary role in the success of online manipulative designs that negatively impact behaviors and judgements. In terms of online choices related to one's personal data, Waldman [455] explains the main pervasive cognitive barrier as anchoring, hyperbolic discounting, loss-gain framing, and over choice. Anchoring makes people over-rely on available information when making their choices (e.g. what other people do), instead of basing their choices on the actual relevance of that information for the situation at hand. Hyperbolic discounting entails the overestimation of the immediate benefits of a certain action, while underestimating its future consequences – this is why people accept extensive digital tracking against free internet content. In loss-gain framing only the positive effects of a certain action are provided or highlighted, while the negative ones are glossed over or vice versa — for example in cookie banners this tactic is often employed to nudge people to consent to personalized advertisement [387, 226]. Lastly, in over choice the excessive number of choices overwhelms and paralyzes users, for instance in terms of mobile application permissions and cookie installation, instead of enhancing their autonomous decision-making.

In their literature review of nudges used in privacy and security for more or less praiseworthy objectives, Acquisti et al. [3] identify additional hurdles. First, loss aversion, which causes people to value the personal information they have lost less in comparison to the information they still have (and thus resist losing it). Second, optimism bias leads people to take unjustified privacy risks based on wrong estimations of their chances of undergoing a negative event. Third, status quo bias which pushes people to stick with default options, like privacy-invasive pre-ticked boxes.

Concerning purchases and decisions on e-commerce services, many cognitive biases have been identified as having a role in transactional decisions, such as overconfidence, present bias and loss aversion [132]. Mathur et al. [296] add to the list more cognitive vulnerabilities. The bandwagon effect refers to when people value something more because others value it; the scarcity bias pushes people to value more what is in short supply; and according to the sunk cost fallacy, people continue with a course of action if they have invested resources into it, even when it is not reasonably worthwhile. Somehow related to this is the restraint bias, which designates people's tendency to overestimate their capacity for impulse control. This is a widely studied effect in addiction settings [216] and plays a role in purchasing and time spending decisions.

Priming effects on purchase decisions

Priming is the effect of stimulus exposure to the response of a later stimulus [33, 209]. Research has demonstrated that the perception of a subsequent stimulus, known as the target, can be influenced by a preceding stimulus, referred to as the prime, even when the prime is visually masked to reduce its visibility or presented very briefly before the main stimulus [53]. However, even when the priming stimulus and the target are separated for a longer period of time and are thus distinguished, the prime is not necessarily processed in an active way [77]. That means, for example, that primes can go unnoticed, but be effective in terms of directing outcomes [147].

The concept of priming has been at the center of much research and practice in product advertising [467, 469, 468], where it can be considered as an instrument of persuasion [464] which is happening outside of the individual's conscious control. Petticrew et al. [359] and Costello et al. [99] demonstrate the use of the priming effect to implement 'dark nudges', i.e., possibly manipulative primes, which could direct users' actions in predetermined directions.

In the domains of e-commerce and online advertising, it has been proven that numeric and semantic primes can have a large impact on the customer's willingness to pay [109]. People mostly interact with e-commerce applications via screens and interfaces, the use of colour, form and other visual-oriented primes [219, 229, 53] in the choice architecture of interfaces can direct the customers' choice towards certain products. The typical tasks performed on e-commerce websites include primarily the visual evaluation of a product through photo galleries and carousels and easily accessible paying options available in a specific, visible compartment. Subliminal priming effectively affects the immediate decision of the user only when the purchase decision will directly follow the persuasive attempt [131], which can be easily engineered in the choice architecture of a digital application.

Functional states of users

The state of individuals affects task performance in complex computer-based systems [299]. Studies show that under conditions of stress, fatigue, and boredom, a deterioration in performance quality and an increased level of errors can be expected [224, 371, 102, 178].

In the context of online deceptive designs, the system can be purposely designed as a responsive agent to these states. For example, several deceptive game patterns operate under the users' state of boredom [78] and even induce it (as in the "Pay to Skip" dark pattern) [473]. User's mental fatigue is also exploited by the so-called 'sneaking' types of patterns [169], which profit from lower levels of attention and cognitive functioning to impose additional purchases of goods or services on unwitting users. Several deceptive patterns that play on a sense of urgency such as fake countdown timers or scarcity such as fake limited availability of a product or offer can also be attributed to the exploitation of these functional states, because their primary goal is, at least partially, to induce stress in users [9] and nudge then to take fast, suboptimal decisions [9], like impulse buying [316].

Other personal conditions

A majority of studies in manipulative design that focus on users have looked at effects on behaviour caused by the existence of different UI elements [38, 45, 280, 111]. However, very few of them have explored how user characteristics influence the experience when facing manipulative designs, like age [15, 45, 447], education [45, 280] or socio-economic status [45, 280]. Some preliminary studies hint towards the need of exploring age as a condition of vulnerability [15, 45, 447, 382, 383]. However, one cannot only look at age as a driver, but at other elements associated with that age. For example, childhood, adolescence, early youth, maternity and paternity, or retirement, embody situational aspects that can contribute to the experience of vulnerability beyond the age.

Children may be particularly prone to fall prey to manipulative designs due to factors such as their immature executive function, susceptibility to rewards, unfamiliarity with data privacy and lack of understanding of virtual currencies [341]. In this regard, prior studies show that children may be particularly exposed to manipulative design patterns due to the massive targeting in in-app advertisement that they are subject to [305, 367] and to the deceptive strategies that are widespread in online games (i.e., loot boxes) [146].

Similarly, for teenagers, there is a social component in their interaction with manipulative designs: they are more exposed to them and tempted to interact with them because of their social relations with family and friends [382]. Sánchez Chamorro et al. [382] resorted to an utilitarian approach to explain why adults are susceptible to some manipulative designs like scarcity cues: adult users buy an item because there is only one good left, and they need it. However, teenagers may be aware of the fake scarcity, but still buy because they would otherwise miss the opportunity of belonging to a group, such as the group that has some cosmetics in a particular videogame. While there is still little research in this regard, among teenagers the utilitarian aspect seems to disappear to leave room for a social one. This social aspect may thus be a specific driver of vulnerability, that is not necessarily only present in teenagers but may be dominant in that age group and further aggravated by other conditions.

On the other hand, older adults may also be exposed to increased risk online, although the evidence in this regard is contrasting. van Nimwegen and de Wit [447] used an experimental setting and found a negative relation between age and falling for some manipulative designs: the younger the user, the more likely to fall into some deceptive designs like sneaking products into users' basket and the use of emotions. Conversely, via a survey study, Bongard-Blanchy et al. [45] showed that older participants had more difficulties identifying manipulative designs. Similarly, [15] found participants above 35 years old more vulnerable because of their lack of acquaintance with the online travel agencies, which was the context studied.

The difficulty of reaching general conclusions is probably because older age is associated to other conditions, such as disability, mental impairment, and digital exclusion, that can exhibit great variation from one person to the other [340]. For instance, the OECD has shown concern about how users with visual impairments can be affected by manipulative designs [340] and recent evidence shows that there can be indeed a troubling effect of deceptive graphical elements combined with poor design practices that do not ensure accessibility, such

as low contrast of refuse links in cookie banners [91]. Even the UK's Competition and Markets Authority (CMA) [132] warns that personal traits (such as age, wealth, and health) as well as temporary situational characteristics (such as unemployment or experiencing scarcity of time, money and social connection) can exacerbate the susceptibility to being influenced by online design choices and can make people obtain worse outcomes from their digital experiences.

6.6.4 Examples of Layered Vulnerability

In the following, we gather some examples that illustrate how certain design elements of digital services can pose a threat to anyone, but are particularly risky for certain ones among their users, often due to a combination of factors that represent the (sometimes situational) layers of vulnerability [283, 290] that people have.

Manipulative designs can increase risks for victims of intimate partner violence. Penzey-Moog [358] describes the real case of a fitness app that made visible their location with any other user nearby by default with the intention of increasing sociability among its users, such as exchanging with other athletes and enhancing the motivation to perform. However, such a social feature carries a great risk. Location sharing can reveal where a person lives and the route they take regularly for exercising, including weather that is a dark or secluded place. This information can be misused by anyone, but it becomes particularly dangerous in the case of survivors of domestic violence that run away from their abusers. Privacy invasive default settings may effortlessly vanish all the extreme efforts that survivors have made to stay alive. Given that privacy settings are often difficult to find, reach or decipher, which default has been selected for the user is not immediately visible nor easily editable, especially when data (e.g., location) sharing happens in an invisible manner, so the user may become aware of the risk too late or never at all. Although privacy risks can be generalized, for survivors of domestic abuse the effect of a privacy invasive default carries an additional danger to their physical safety.

Manipulative designs can increase risks for LGBTIQ+ community. Another example of how the design of privacy settings may excessively affect certain communities is given in Sannon and Forte [385]. Privacy settings that are difficult to find and manage are costly especially for those users that need to frequently switch between identities, for example because they are forced to conceal or separate certain aspects of their lives (e.g., members of the LGBTIQ+ community, political opponents in autocratic regimes, etc.). The ability to adjust privacy preferences often requires a certain level of digital skills, which is often associated with one's socio-economic status, as recalled earlier. Data management may be further complicated by language barriers and the financial cost of connectivity in certain geographical areas. Failure to protect one's own privacy due to design barriers may result in serious implications on the lives and welfare of such communities.

Manipulative designs can increase risks for users with mental disorders. The UK's Money and Mental Health Policy Institute [204] denounce that many online gambling websites contain easy manners to make deposits while they conceive cumbersome, frustrating processes to withdraw funds. Moreover, tools that help setting spending limits and imposing self-exclusion

are often hidden or ineffective. Although all users are exposed to such tricks and the nefarious (financial, emotional, relational) consequences of gambling may fall on anyone, people suffering from mental health issues are particularly vulnerable, due to a mix of factors. Common symptoms of such problems are impulsivity and low problem-solving ability which are associated with low control and difficult decision-making, including risk assessment [204]. When coupled with designs that nudge people to continue playing and spend money, the mix is lethal. The problem is exacerbated by online advertising and marketing strategies that are impossible to avoid [204] and lure users into gambling websites by targeting specifically those vulnerabilities.

Manipulative designs can increase risks for individuals with a predisposition to impulse buying. Internal conditions, as predisposition to impulse buying, can be accentuated, and even targeted, by manipulative designs. Moser et al. [316] carried out a study on dark patterns used on the top US e-commerce websites, noticing that they often include features that aggressively encourage impulse buying, such as lowering the perceived risk of carrying out transactions online, leveraging social influence and enhancing the perceived local and temporal proximity to the product. The study shows that such practices counter the will and capacities of consumers who desire to curb their impulse buying. What is more is that those that find it difficult to make ends meet experience a more severe detriment of such commercial practices [132].

6.7 Discussion: Towards a Multidisciplinary Assessment of Vulnerability

6.7.1 Building Resilience Against Vulnerability

Throughout this article we have argued that deceptive design patterns may exert their effects on all users of digital technologies but can also target specific vulnerabilities and thereby be more detrimental to certain individuals or communities. Policy, practice, and research need to assess, and ultimately address, such challenges. It is hence important to discuss the implications of our arguments for building the resilience of individuals and groups against deceptive designs. Since human beings are embedded in a common social, economic, institutional, and legal fabric, they are by nature interdependent. Their dependence on external conditions is the concrete manifestation of human vulnerability, which is universally present. This is why a state of non-vulnerability is impossible to achieve and should hence not be the goal of policies. Rather, resilience [138] should be promoted, namely the set of all those physical, human, environmental, and social resources that make human agency possible and enable people to cope with the implications of their inescapable vulnerable dimension. Resilience to dark patterns should not only be promoted as an ex-post mitigation mechanism, but also as an ex-ante solution. Many different actors of the digital economy can supplement the necessary resources to prevent the adverse effects of online manipulation.

6.7.2 Practical Hurdles to Risk Assessment Feasibility

With its purpose of proactive mitigation of the many drivers of vulnerability, risk-based regulation seems fitting to approach the ever-evolving reality of deceptive design patterns in existing and emerging technologies. Whereas the domain of data protection can count on an established body of risk assessment methods, the implementation of the systemic risk assessment envisioned in the DSA is, at date, still to be defined. Regarding

AI systems, the risk appraisal frameworks that are being developed are several and sometimes divergent [465]. Such an uncertainty leaves organizations unsure of how to plan and operate their development and commercialization activities in a safe and compliant manner, especially when it comes to implementing safeguards against digital manipulation, which is a young area of research and practice.

Further, if the issue is not addressed on a systemic level, individual safeguards on individual technologies risk to be ineffective. For instance, market-related macro-conditions foster the global tendency of ad personalization based on massive collection of personal information, while the competition for users' attention gave rise to a race to the bottom with the subsequent proliferation of dark patterns on the overwhelming majority of online services. However, macro conditions are often hard to account for in application-specific risk assessments.

Moreover, when risks are highly contextual, it is inaccurate to only ascribe them to a static set of functionalities of a certain technology (as a rigid reading of the DSA would propose) or to the purpose and domain of use of the technology (as the high-risk categorization of the AI Act suggests). In AI systems, moreover, deception is sometimes a mere strategy for a pleasant, functional interaction design. But it is because of this reason that, if the risks are not counterbalanced or removed ex-ante, these technologies can also produce greater harm than those that have been identified as harmful a priori. Given that the elements of vulnerability are many and intertwined, continuous assessments may result in a laudable theoretical effort, but be excessively costly and clash with the priorities, needs and timelines of business practice.

Such challenges risk jeopardizing the feasibility of risk assessment. A standardized approach would undoubtedly simplify the procedure and, supposedly, be experienced by organizations as less of a burden. Going beyond mere compliance reasons, technology developers need to be persuaded of the motivations why an accurate, reliable risk assessment can constitute a helpful tool. In addition, there is the need to develop a comprehensive methodology that includes the evaluation of the risks entailed by all sorts of digital manipulation within other kinds of risks, since it would prove burdensome to carry out separate ad hoc assessments. Accounting for and understanding the internal constraints for companies and actors involved in the digital architecture when applying risk assessments is a first necessary step, and potential future work we envision.

It is paramount that policymakers as well as civil society take part in the establishment of the methodologies for such an appraisal and the appropriate mitigation measures. Articles 34 and 35 of the DSA, for example, already foresee such a participatory approach, wherein the European Commission as well as the European Board for Digital Services play an active role

in the oversight and the recommendation of measures for the management of systemic risks of very large online platforms. A similar role will be taken on by competent authorities in the activities carried out within the regulatory sandboxes established by Article 57 of the AI Act, that will offer controlled environments that enable the testing and validation of innovative AI systems before their placing on the market. Regulatory sandboxes will also provide the possibility to identify risks upfront and devise timely and effective mitigation measures.

6.7.3 The Role of Empirical Evidence for Transdisciplinary Action

Existing empirical studies can be helpful in determining a broad range of potential threats that need to be factored in the risk appraisal methods. However, such an approach falls short of being complete, as it excludes all those risks and factors that have not been examined yet. The OECD [340], for instance, observes that most research studies so far have only examined internal drivers of vulnerability (e.g., age, socioeconomic status) while neglecting other relevant elements. There may be additional factors of vulnerability, though, that are difficult to identify, quantify, correlate and report (either directly or indirectly), since they relate to the experience of users and not to an observable change in behaviour. For example, inter-partner abuse victims may pertain to any socioeconomic class and experience privacy harms differently, but it is hard to quantify them in an observable way. Similarly, older adults are not a homogeneous group, and there is not an specific age threshold after which one should be considered an older adult [29]. Therefore, there is not an age in which users automatically become more vulnerable [340], and context plays again an important role. The fact that people can be vulnerable in one situation but not in another makes it cumbersome to exactly determine the drivers of vulnerability.

Social sciences methods and computational methods that collect empirical evidence are the necessary candidates to bridge this gap, as suggested indeed in Recital 90 of the DSA. As pointed out by Gray et al. [171], there is a pressing need for transdisciplinary approaches and knowledge transfer to understand and fight the effects of manipulative designs. For example, design scholarship can help regulators understand the impact of technology design on vulnerabilities.

To broaden the understanding of this multifaceted reality created by digital markets, it is crucial to run user studies that include participants other than highly educated populations in developed countries with digital access to survey platforms (e.g., elderly, kids, teenagers, or low-educated people), as well as organize research designs in contexts that can expose vulnerability. In the same way, looking at the experience of interacting with manipulative designs, and not only at the effects in behaviour that design features have, will help to disentangle the contextuality and situatedness of vulnerability to deceptive design patterns. Hence, it is not only about “what makes users vulnerable”, but also the “how and why”.

Contacting and collaborating with organisations such as NGOs that work with specific populations in real-world contexts can be the first necessary step to carry out this endeavor with all the necessary ethical considerations O'Brien et al. [348]. The engagement of those that have firsthand experience and knowledge of certain realities is gaining importance for risk

assessment. For instance, to determine the dangers that technology can cause to victims of intimate partner violence, Slupska and Tanczer [415] propose to involve affected groups and communities in the traditional threat assessment methodology: they can help mapping out the actual threats and devising measures for those types of harms that are less tangible than financial losses. The engagement with various stakeholders and the evidence that empirical science can thus provide can support realistic risk assessment and nurture good practices.

6.7.4 Implementing Fairness and Fair Design Patterns

Across various domains, there are proposals for a “fairness-by-design” duty Siciliani et al. [405] that could even be incorporated in the UCPD revision as a general obligation for businesses so that “products, user interfaces and commercial communications [...] be designed in a fair manner” [346](p. 13). Fairness is more daring and more encompassing than the principle of transparency that has the goal of disclosing how a system or process works. The multidisciplinary community of researchers, regulators, civil rights defenders and businesses who work to contrast digital deception has so far mainly proposed transparency-enhancing measures (see e.g., [43]), which are necessary but not sufficient to fight dark patterns. It is now time to define and apply fair design practices as well as incentives for their adoption and determine their fitness for protecting vulnerable people and increasing their resilience. The community has achieved astonishing results in the identification and exposure of problematic design practices, for instance through studies aimed at detecting dark patterns automatically [235, 192] or at demonstrating their influence on people’s decision-making (see Section 6.6 and for an overview [171]). However, in the already highly regulated digital sphere, there is the urgent need to successfully promote fair design patterns that can be adopted easily and safely by businesses. It is paramount that the high-level (often intertwined) requirements provided by existing and upcoming regulations are translated into simple, operational instructions, accompanied by examples of good practices, that designers of digital experiences can effortlessly understand and apply.

Proposing one-size-fits-all solutions is not the goal, because there is often a subtle distinction between design patterns that are legitimate and appropriate within a certain context and those that are not. Rather, there needs to be an inventory of good design practices that one can draw from, in combination with knowledge about the harms that prospect users could suffer and about the methods for assessing, mitigating and eliminating risks. For what concerns privacy policies, for instance, there exist libraries of design patterns that collect, organize and make readily available good practices [374], such as the French Data Protection Authority’s library on transparency-enhancing design patterns (CNIL) [110]. Even though it is for now easier to find and copy-paste bad practices since they are so widespread and the incentives for their adoption are high, designers and developers should be empowered to reuse fair design patterns and adapt them to their specific contexts.

Designers can also play a crucial role in the implementation of fair design patterns as they can embed awareness to vulnerability factors within the design process, thanks to a wide range of methods through which they can evaluate their work’s potential impact. The use of

personas, understood as a sort of average user that the planned design would target, and anti-personas, as those users that might be excluded from the planned design, can help to assess the impact of the adoption of a certain design element on a diversified variety of users. Similarly, specific methodologies and toolkits enable the impact assessment for inclusive design and ethical design [206, 302], thereby supporting the development of less harmful designs.

Moreover, businesses can empower designers in their decisions [462, 379, 83]. Fostering governance models in organizations where legal departments, decision-makers, and designers work together within a check-and-balance system can help to broaden the adoption of ethical and safe interface designs. Without changing the market conditions and without developing economic incentives for adopting vulnerability-aware fair design patterns, this paradigm shift will not happen. The Digital Markets Act is a first, important step in this direction and, indeed, it contains provisions against dark patterns. Additionally, the Data Governance Act and the fair data economy it aims to foster can promote digital services that are safe and respectful by design: data intermediaries, data cooperatives and data altruistic organizations have the precious opportunity to conceive and design experiences for data sharing and consent that are in stark contrast with the deceptive status quo of data-hungry digital services.

6.7.5 Conclusions

Counting on the multidisciplinary expertise of the authors and the trans-disciplinary knowledge that was generated through their collaboration, in this article we have argued that the harmful deceptive design patterns in digital services can be more detrimental to certain individuals or communities due to macro, meso and micro conditions. In the age of service personalization and hypernudging, there is the risk that manipulative attempts will increasingly be able to exploit such vulnerabilities to strengthen their effectiveness and weaken people's resilience even more. Risk assessment is becoming inescapable to account for the ever-evolving nature of digital technologies and the vulnerabilities they engender, but there are open questions on how to carry it out in a reliable and practicable manner. All actors of digital markets need to be involved, and held accountable when appropriate, in the creation of fair-by-design experiences.

6.8 Chapter Takeways

This chapter provides a theoretical analysis of what vulnerability to manipulative designs means in legal texts and how with HCI theory these concepts can be better nuanced. Here there are two fundamental problems revealed: vulnerability to manipulative designs in legal text is not clearly defined and, therefore, hard to operationalise by practitioners when they want to evaluate technologies. Thus, while here we refer to "risk assessments" to assess technology, this is not a tool only pertaining to the legal scholars, but one that designers also contribute to. By using HCI theories and tools, this chapter proves useful to investigate the situated experience of users with manipulative designs in order to understand their vulnerability.

Part III

Contextual Drivers of Vulnerability



This part focuses on the **exploration of experiences that lead to vulnerability** to manipulative designs.

As explained in **Chapter 2** vulnerability has to be driven. This part shows three studies with different populations that investigate their experiences with manipulative designs and that allow the extraction of experiential drivers of vulnerability.

Chapter 7 focuses on teenagers' experiences in three contexts: video games, social media and e-commerce. It emphasises the social aspect of manipulative designs and provides contextual harms tied to this population.

Chapter 8 uses magic machines workshops to understand the experiences of manipulation in older adults. It showcases their needs when it comes to resisting manipulation.

Lastly, **Chapter 9** focuses on users with lower levels of digital skills and their experiences with manipulative designs. It explains how the imaginaries of manipulative designs are related to the different ways users deal with them. In every chapter, different contextual drivers of vulnerability, as well as their corresponding challenges to design counter-measures, are provided.

“My Mother Told Me These Things are Always Fake” - Understanding Teenagers’ Experiences with Manipulative Designs

This chapter is based on the following peer-reviewed publication:

Lorena Sanchez Chamorro, Carine Lallemand, and Colin M. Gray. 2024. “My Mother Told Me These Things are Always Fake” - Understanding Teenagers’ Experiences with Manipulative Designs. In Proceedings of the 2024 ACM Designing Interactive Systems Conference (DIS ’24). Association for Computing Machinery, New York, NY, USA, 1469–1482. <https://doi.org/10.1145/3643834.3660704>

Abstract. Manipulative and deceptive design practices are ubiquitous, impacting technology users in various ways across several domains. Certain groups are likely more susceptible to these impacts but have not received sufficient attention yet. In this paper, we seek to characterize one such understudied group, describing teenagers’ experience of manipulative design. We conducted semi-structured interviews with six teenagers between 15 and 17 years old, to understand their daily interactions with manipulative designs in three contexts: social networks, video games, and e-commerce. Using reflexive thematic analysis, our findings describe how risk is a shared experience for teenagers, and interrogate how teenagers’ personal and social context shape their experience of risk. We relate our findings to existing knowledge about how the general population is impacted by manipulative design practices and consider opportunities to further understand and support the experiences of teenagers and other vulnerable groups

7.1 Introduction

"Please stop using your phone at dinner! Do you know what time it is? You have school tomorrow, turn off the computer! Stop playing videogames, go outside and play with your friends." Manipulative designs—commonly known as “dark patterns”¹ — are “user interface design choices that benefit an online service by coercing, steering, or deceiving users into making decisions that, if fully informed and capable of selecting alternatives, they might not make” [296]. These manipulative practices have pervaded the lives of teenagers [341, 180, 111], who are considered a vulnerable population online by researchers [273, 461] and policymakers [41, 341, 340], as evidenced by the 2023 OECD report of online vulnerability in consumer protection [340]. Teenagers' specific position and understanding of the online world makes them more willing to take risks while lacking resources to cope with harm [218]. Despite the impact of manipulative designs on teenagers, which is a rising concern [171], there is still an important research gap.

In the past years, scholars have contributed to a growing body of research on manipulative design, with particular emphasis on studies that evaluate the existence of manipulative designs and seek to understand users' perceptions and behaviours in different contexts [171, 78, 45, 280, 297, 336]. The pervasiveness of manipulative designs calls for expanding this body of research to a variety of populations, which in turn will support researchers and practitioners in designing suitable interventions for users.

We aim to start a conversation about teenagers' specific needs when fighting manipulative designs by understanding their everyday experience with these designs and documenting what is unique in their ecologies of use, understood as the different structures in the environment that surrounds them [57], which might impact their experience. In collaboration with an NGO working with families at risk of social exclusion, we interviewed six teenagers to understand their relationship with manipulative designs in three scenarios: video games, social networks, and e-commerce.

Our paper makes several contributions to HCI research. To the best of our knowledge, we document the first study that focuses on teenagers' experiences with manipulative designs in their daily interactions with technology, providing insights into how their ecologies differ from the ones of adults and by explaining the effects of teenagers' environment on their relationship with manipulative designs. Our findings represent a starting point for understanding the ecologies of manipulative design in teenagers, including their potential position of vulnerability, offering a socially-focused solution space to prevent the effects of manipulative designs. By bringing an experiential perspective into the realm of manipulative designs, we aim to support designers by providing new challenges to design countermeasures to these designs. We expect to inspire further research focusing on vulnerabilities online.

¹The research community is studying this phenomenon using a variety of labels, including deceptive design, nudges, anti-patterns, and most dominantly, “dark patterns.” Following the ACM recommendations on diversity and inclusion [143] we hereby use the term “manipulative designs” to describe this phenomenon.

7.2 Related Work

7.2.1 Online Vulnerability and Manipulative Design

Scholars in the legal domain have argued that users are in a state of vulnerability online, and the consequences of this vulnerability [194, 340, 290]. Vulnerability is a multilayered construct [283] and translates into users finding themselves in a position of susceptibility towards technology, where the impact of online threats is amplified [290]. Thus, researchers have considered teenagers a special category of vulnerable population because of their risk-taking behaviors and fewer defenses to cope with potential harm [461, 340]. While we embrace the critiques towards “vulnerable” as a user category by default [283], we believe that studying this population might be a starting point to debunk assumptions and disentangle factors of online vulnerability that consequently allow us to detach it from categories of users. Some preliminary studies are pointing out socio-digital vulnerability as the idea that the environment puts users in a more susceptible position [113].

Concerns regarding harm and agency have also reached the field of manipulative design [341, 340, 41, 14]. Manipulative designs are associated with various risks, including psychological harms such as emotional distress and cognitive burden, alongside autonomy loss, financial harms, or privacy-related harms [297, 181]. As Bongard-Blanchy et al. reported [45], even the users who can identify manipulative designs remain unsure about the impact these designs might cause. For teenagers, risks online can be further described in terms of content, contact, conduct, and commerce—known as the 4C framework—which can be induced by manipulative designs [140]. Indeed, in a recent systematic literature review on teenagers’ harms caused by social media, Sala et al. [377] showed how some design elements — e.g., “like mechanisms” — are associated with some emotional harms. This harm-based approach becomes even more relevant in light of new regulatory frameworks that ask companies and designers to assess the impacts of their designs; for instance, Article 34 of the EU Digital Services Act includes assessing health impacts or negative effects on minors from the platform.

Manipulative designs are complex given their ubiquity and subtle mechanisms [379], with attributes that have been defined as coercive, manipulative, deceptive, and steering in ways that users would not intentionally desire [296]. Manipulative designs have also been discussed in relation to different theories of digital nudges, sludges, and online manipulation [379, 304]. The subtlety of the mechanisms makes them hard to perceive by users; therefore, this study takes a relational approach to understand the relationship between user and manipulative design, accounting for users’ felt online manipulation as a proxy, as already used in Gray et al. [163]. We explain this rationale through the idea of the relationality of manipulative designs. Borrowing Star’s terms [424], manipulative designs are relational: the only way users perceive them is when an interaction leads to a negative outcome. Gray, Kou, et al. [169] anticipated this idea through Norman’s gulfs [335]: manipulative designs are perceived as a mismatch between what users expect from the interaction and what they receive. Similarly, Gray, Chen, et al. [163] reflected on the idea of “temporality of manipulative designs,” supporting our goal of investigating how teenagers experience that manipulation on

an ecological level over time.

7.2.2 Experience of Manipulative Design Practices

Research on manipulative design has adopted different methods, audiences, and contexts [171]. Among user studies, the effects of behaviour change by different UI elements have been studied in the domain of cookie banners and privacy [442, 162, 34, 247, 46], streaming platforms [78], digital services [45], and social media [308]. Increasingly, manipulative design practices are not only experienced by users via specific targeted UI elements, but also as part of a user journey, system, or service delivery strategy [172]. Thus, as part of our study framing, we focus on the digital systems that users reported engaging with, while also recognizing that the larger systems and ecologies these systems are embedded within are large and complex, and are driven by different motivations (e.g., the “attention economy” or “influencer economy”).

Few of these existing studies have explored how user characteristics influence the experience of manipulative designs, such as education [45, 280] or age [15, 447, 383]. In this section, we describe the main findings from previous studies that help to understand the relationship between users and manipulative designs in different contexts. While some might include young adults, none of them included teenagers.

A limited set of the literature focuses on the experiential aspect of manipulative designs. A survey conducted by Bongard-Blanchy et al. [45] showed a relationship between people's perceived self-efficacy in resisting these designs and their capacity to recognise them. Maier and Harr [287] showed undergraduate students examples of manipulative designs and their definitions, explaining how their perceptions depended on the perceived harm resulting from the designs. Gray et al. [163], resonating with Avolicino et al. [15], additionally explored the ranges of negative emotions users experience after realising the manipulation, including: distress, upset, guilt, fear, hostility, irritability, shame, and nervousness.

Experiences with attention capture deceptive patterns—deceptive patterns that impact users' attention spans—have been described on video platforms [315, 281, 78]. For instance, Lukoff [281] related these patterns with the sense of agency online. Non-consent mechanisms feel deceptive to users, who view these mechanisms as triggers of pressure to spend more time on the platform; these mechanisms include disliked ads that pop up, accidental clicks by ads, or autoplay turned on unnoticed. Chaudhary [78] also highlighted the importance of a ‘mindlessness’ (p.9) experience while interacting with the platform

Research on video games has explored the effects and experiences of some manipulative designs on young consumers. Loot boxes, namely features inside the videogame that provides a service or digital good with a pre-set probability determined by an algorithm, are a concern because of their relationship with potential gambling disorders [177]—which can include manipulative design patterns, as described by Zagal et al. [473, 74]. Pay-to-win and in-purchase games features have also been discussed as part of the players' self-development, socialisation, and identity within a community [150, 65, 271]. Thus, in the field of psychology, the literature aiming to establish the relationship between problematic gaming and gambling

disorders to these design systems is growing [122, 159, 475]. Still, from an experience design perspective, Dechant et al. [106] call for debunking vulnerability in videogames by exploring the harms that come from design rather than focusing on users, which aligns with the purpose of this study of exploring design with the experience of harms.

In e-commerce, Moser et al. [316] explained impulse buying through manipulative designs, and how buyers reclaimed more friction to help them reduce impulsivity. Low stock messages, hidden costs, and aesthetic manipulation can influence the users while buying online, reducing their agency [15]. In recognising these features, van Nimwegen et al. [447] found that younger users have difficulty identifying ‘Sneak into basket’ designs, positing a relationship between the perceived honesty of the website and perceived good navigation. Additionally, Fear-of-Missing-Out (“FOMO”) is triggered by limited offers in purchases [433] but is also a rationale for giving up on privacy. Thus, users might accept settings that they do not want to, despite recognising the bad design or experience, because they want to belong to something; this is what Westin and Chiasson [458] call “participatory reluctance” using Casiddys’ theories (as cited in [458]).

To the best of our knowledge, little to none of this research focuses specifically on teenagers’ experiences with manipulative designs. Exploring teenagers’ experiences and ecologies is necessary to understand their vulnerability and interaction with manipulative designs. Therefore, this study addresses the following research question: *How do teenagers experience manipulative designs during online interactions?* Understanding the unique factors of teenagers’ experiences will give the HCI community a more detailed foundation for designing interventions to protect teenagers.

7.3 Methodology

7.3.1 Participants

Participants were recruited in collaboration with an NGO conducting socio-cultural interventions with populations at risk of social exclusion in Madrid (Spain). We opened the call for participation to all teenagers who regularly attended the activities of the NGO. The consent form was communicated to the teenagers and their families through the NGO a month earlier to give them time to read the information and formulate questions. Those teenagers whose parents and themselves brought a signed consent form could participate. This study received ethical approval from the University of Luxembourg.

Our participants’ involvement in a socio-cultural organisation might suggest a higher acquaintance and awareness of manipulative elements than teenagers not being part of such a community. Noteworthy, the teenagers we interviewed did not receive any education on digital skills through this organisation; the activities the NGO conducted supported them in their homework and provided a space to spend some healthy leisure time.

In this study, we rely on an interpretative approach, aiming to unfold a phenomenon and to get a rich understanding of a small sample of participants’ lived experiences. We do not

make an epistemological commitment that focuses on either generalizability or representativity [418]. We describe participants' profiles to better understand their backgrounds and contexts. We assigned them pseudonyms to preserve confidentiality.

- Ineke was born in 2007 in Spain, but her family comes from Morocco, so her mother tongues are Spanish and Arabic. She identifies as female. She has two devices to connect to the internet, and the first time she did it she was 9 years old. She now uses the internet more than once a week to search for information for school, play some video games, and use TikTok, but she does not declare herself a big fan of technology. She is in the fourth year out of four of mandatory secondary education ("ESO"). She has a little brother, who she spends a lot of time playing with, and a sister. She lives with her father who is a house janitor and her mother who, in the words of the participant, "does not work." We discussed social media, video games and e-commerce websites.
- Lola was born in 2006 in Peru; her mother tongue is Spanish. She identifies as female. She has two devices to connect to the internet at home, and she was 7 the first time she used the internet. Nowadays, she spends more than 4 hours a day online. She is in the fourth year out of four of mandatory secondary education ("ESO"). She lives with her mother, who is a technical nurse, and her younger sister. She loves video games and spends hours playing with, in terms of the participant, "online friends." We talked about video games and social media but did not have time to discuss e-commerce platforms.
- Alex was born in 2007 in Peru, his mother tongue is Spanish. He identifies as male. He has three devices to connect at home and uses the internet for more than two hours a day. He started to use the internet when he was around 10 or 11 years old. He watches movies and shows on Netflix or from pirate websites. He loves video games and plays a lot on his mobile phone, where he particularly enjoys horror mobile games - although sometimes he is so scared of them that he prefers to play in 'safer' modes. He normally watches a lot of shows on the Internet as well. He lives with his mother, who is a nurse auxiliary. We discussed video games, social networks, and e-commerce platforms.
- Oskar was born in 2005 in Spain and identifies as male. His mother tongue is Spanish. He uses the internet more than two hours a day, and normally he does it to play video games, use social networks, and listen to music on Spotify. He has three devices to connect to the Internet at home, and the first time he did it he was around 7 to 8 years old. He is in the first year of Bachillerato out two. And his parents work as a dressmaker and delivery man. He chose to speak first about e-commerce because he never had the opportunity to discuss it, and later on, video games, and he finally stopped the conversation before talking about social networks.
- Victor was born in 2007 in Spain and identifies as a male. His mother tongue is Spanish. He uses the internet more than four hours a day, for school-related tasks, but also for playing video games. He loves video games, especially platform fighter ones. He has two devices to connect to the internet at home, and the first time he used the internet,

he was 7 years old. He is in the third year of high school and lives with his mother, who works in a kitchen. He enthusiastically decided to start the discussion with video games, then we talked about social media, and lastly, we discussed e-commerce.

- Anna was born in 2005 in Spain. She identifies as female and is in her senior year of high school. Her mother tongues are Spanish and Arabic. She has three devices to connect to the internet at home, and she started using the internet when she was 10. She normally uses the internet more than four hours a day, mainly for studying, watching stuff, social media, and online shopping, which she identifies as her weakness. She loves shopping and “could spend hours and hours just looking at clothes on the internet.” She has a brother who often plays video games on the PlayStation and on the mobile phone. Her father works in a fruit shop, and her mother in a SME. We discussed e-commerce websites and social media. We did not have time enough to go in-depth into video games, although she gave us some main impressions.

7.3.2 Procedure

We conducted six semi-structured interviews with teenagers aged 15 to 17. The interviews were conducted in March and April 2023 on the NGO premises. After the introduction and a reminder on data protection and participants’ rights, we proposed three discussion contexts to participants: video games, social networks, and e-commerce.

The interview guide² included the following topics: use of technology (social networks, video games, and e-commerce), critical incidents with technology, perceived risks, intentional coping strategies, and the role of parents. To make the interview as comfortable as possible, we offered participants to choose their preferred context(s) and to stop at any time.

Each topic consisted of two parts. Based on the critical incidents technique [417], the first part invited participants to think and reflect on the moments in which they had bad experiences and felt deceived or manipulated when navigating online. We did not show any manipulative design at the stage to avoid priming participants. In the second part, participants were shown manipulative designs and asked how they perceived them.

These examples represented three specific contexts that might require different considerations and trade-offs from users [171], which are familiar to teenagers and where manipulative design techniques are common [316, 473, 308]. To select the manipulative designs, we considered the high-level patterns from Gray et al. [172] (See Supplementary material). These were not meant to be exhaustive but to trigger a conversation with participants.

7.3.3 Data Analysis

After transcribing the interviews, we conducted an inductive reflexive thematic analysis [51, 50]. The transcripts were transcribed in Spanish and then translated into English with the help of automated tools so the research team could understand and discuss the themes. The

²The interview guide and protocol are provided as Supplementary Material.

transcriptions were coded with different coding strategies to enrich the analysis [378]. *In vivo* coding takes literal excerpts from participants, which becomes fundamental to give voice in their own terms for our studied population — e.g., “I have to buy it because otherwise I miss it.” *Descriptive* coding summarises the meaning of a specific phenomenon shared by the participants — e.g., *Spending money in the game attaches to the game*. Lastly, *versus* coding looks at the data from the perspective of a dichotomy or concept opposition, which enriched the analysis by providing specific conflicts in the ecologies of teenagers - e.g. Having obligations vs Not having obligations. An example of codes and themes can be found in the Appendix.

With the coded interviews, and using maps created with MAXQDA, we constructed initial themes around main concepts (e.g., family, protection). Through an iterative discussion, we refined the themes and their relationship. Additionally, we drew on theoretical memos written by the first author that explained perceptions of the data, similarities and differences between participants, and connections with theoretical phenomena and literature. When looking at harms, we used a deductive-inductive approach, starting on the harms framework described by Gunawan et al. [181], and extending it inductively. Similarly, when explaining the manipulative designs involved in the experience of participants, we have used existing standardised terminology in the community by using Gray et al. [172]³ ontology. This ontology also gathers “attention capture deceptive patterns” from [315] and bundles it into one category within the meso-level patterns. However, attention capture deceptive patterns are a combination of 11 different design strategies. Thus, when reporting these attention-capture deceptive patterns, we will use the specific terms coined by Monge Roffarello et al. [315].

Noteworthy, during the interview, insights about internet and video game consumption emerged. This data was not deemed relevant to our research when these insights did not overlap with manipulative designs or non-planned actions of the participants caused by those designs, which is the main driver of this study.

7.3.4 Ethical Considerations: Research Through Care

Given the circumstances of our participants, care was our priority in conducting our interviews, justifying specific measures in recruitment, interviewing, and debriefing. To avoid burden on the teenagers, and facilitate participation, we conducted the interviews in the same slots that teenagers go to the NGO activities. However, the priority was to support the activities performed at the NGO, which thus had priority over our interview if the teenagers preferred it. Additionally, we relied on the teenagers' attendance at the NGO.

Although the interviewer was introduced as a local neighbour and researcher, we understood that this could give participants the feeling they were in an imbalanced position and that they ‘had to answer’ everything. For that reason, we put a lot of emphasis on their rights as participants to withdraw and not answer to anything they did not feel comfortable with. Similarly, dividing the interview into thematic blocks would give an easy option for participants to stop if they felt the need.

³There is a set of the literature, especially in the legal domain, that expands financial harms to anticompetitive harms for companies. As our work focuses on the individual approach, we are not considering that type of harm

A debriefing was done with the teenagers at the end of the session, and parents were provided with a copy of the debriefing information. Participants were fairly compensated under the NGO's conditions and provided a brochure about online safety. The brochure can be found in the Supplementary material.

7.3.5 Positionality Statement

For this study, we embraced our position and subjectivity as a resource [161] given the commonalities and differences in our expertise and position towards the present work. The study of manipulative designs and their relationship with vulnerability has a personal drive for the first author of this paper, who shares the socio-economic, cultural background, and neighbourhood with the study participants. The first author recognised themselves in the experiences and stories told by participants — and presented themselves as such during the interviews — but they also acknowledged their current position of privilege being now an outsider of the group. The first author is a frequent user of social media (e.g., Twitter, streaming platforms like YouTube Shorts, e-commerce, and video games), but to get a better first-person perspective of manipulative designs in different contexts, they also used TikTok, Clash Royale, FIFA, Candy Crush, Free Fire, and Fortnite before the interviews. The second author has experience in supervising HCI studies with children and teenagers, mostly in the education and preventive health domains. She does not recognize herself in the experiences of the participants but relates them to family relatives. While she does not share the participants' cultural background, she comes from a similar socioeconomic background. Her close relatives — both adults and children — are likely vulnerable online and often refer to her as a source of information against manipulative designs. The third author has experience in supervising HCI studies with design practitioners, primarily relating to UX and product management, and they have designed digital services in their past work as a designer. They come from a similar socioeconomic background as the participants, but not their cultural background, yet see similar patterns of interaction with their close family members.

7.4 Findings

In this section, we describe teenagers' experience of manipulative design practices, explaining how the environment of teenagers impacts the experience of manipulation through the following themes: (i) risk is a shared experience, (ii) the personal and social context influence the risk experience, and (iii) the (un) conscious experience of harms.

7.4.1 Risk is a Shared Experience

This theme encapsulates the idea that, through their interaction with others, teenagers identify harms coming from manipulative designs enough to develop awareness that leads them to use certain mechanisms, but also engage with risks as part of a social context. The role of

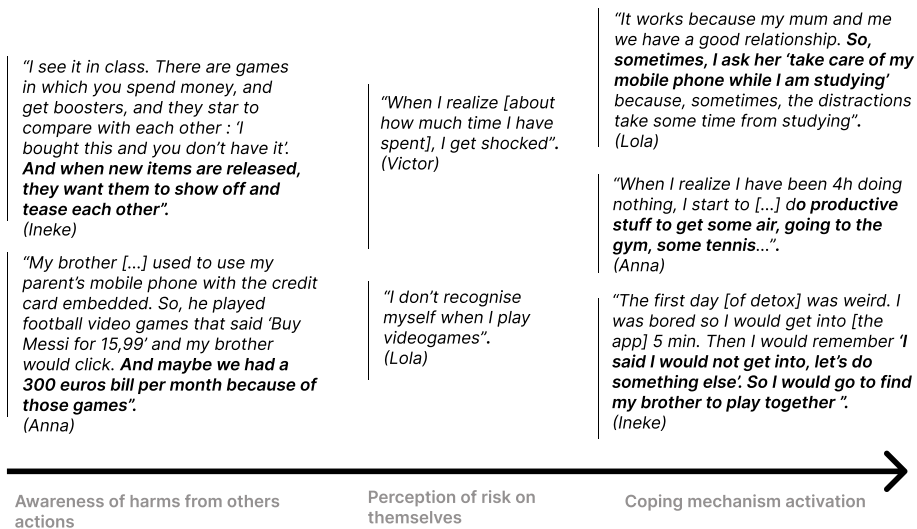


Figure 7.1: Examples of how risk awareness from others' actions moves to self-awareness and coping mechanisms

family and friends in this social approach to risk was a recurrent topic brought up by participants without being prompted. Family and peers helped raise awareness of risks, but peers often induced risks that may lead to harm. Lastly, family plays an important role in teenagers' coping mechanisms.

Seeing harm on others raises awareness

All participants described the negative effects of manipulative design on their peers and family. The effects would sometimes reach them, including scams, deceptive designs, addiction, time waste, money loss, insecurities, depression, and social comparison. For instance, Ineke described how her classmates challenged each other and showed off their new limited skins and objects on video games, incentivising others to buy them as well. Anna also explains how her brother's misuse of online video games caused a loss of money and trouble for the family. Oskar described his distrust of e-commerce websites after his mother was scammed. Similarly, participants have seen friends suffering from the negative effects of "comparing themselves with others" on social media. Participants showed how they are experiencing, being told about, and learning from these risks.

Seeing the effects on others made participants reflect and be aware of some of the risks of manipulative designs. They tried to take action when they started to see similar effects on themselves. Ineke, Anna, and Lola (Figure 7.1) explained how they sometimes needed technology detox, particularly after feeling bad from spending too much time on social media or games. Ineke purposefully stopped using TikTok for 10 days, hiding the app where she could not find it easily. When Anna realized she had spent too much time on her phone, she focused on outside activities or spending time with her family. Both explained how this

happened progressively: *“scrolling, scrolling, and suddenly you spend an hour instead of 30 min.”* Lola had a similar non-purposeful experience. When her phone broke, she realised that *“life offline is more calm.”* Now, when she has something important to achieve, she asks her mother for help to mitigate the risk.

Relationships with others lead to risk

We observed that participants take some risks unconsciously, perhaps motivated by manipulative designs as a part of a shared experience. Ineke shared examples of technology unease with her friends, including an instance where she was not allowed to buy products online because her mum was the victim of a scam. Alex explained how he could only play some video games that cousins would pirate for him without being aware of the privacy risks that pirated products may carry. Oskar also mentioned being tempted to buy more items when pop-ups appeared when being with friends because they teased each other. Some participants explained spending time with their parents through the use of applications — discussing the latest news on social media or helping parents pass a level in a video game. For participants, risks from manipulative designs also came from their internal social dynamics with family and friends.

Although participants associate some positive outcomes with manipulative design practices, they often led to the extension of risks among peers. Victor, for instance, described how social pyramid tactics — i.e., inviting your friends to download mobile games — give your friends a new game to discover without realising the impact on privacy. When he was 12 years old, Oskar would waste his small savings on skins cosmetics on videogames just to *“feel cool”* among his friends. Both Victor and Lola described limited skin offers as something worth paying for. As Victor explains (Figure 7.2), limited purchase time and scarcity of skins on games are perceived as exclusive. Owning these items meant that the players were part of a specific moment in the game (e.g., season, battle pass), which provided heightened status and a feeling of belongingness. It is their relationship with others and the feelings among others that make that scarcity special and worth it.

Family helps me to cope

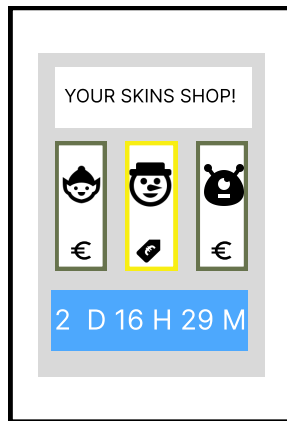
Participants reported how their parents supported them in establishing coping mechanisms when they asked for help. Their parents warned them about the abusive use of social media or video games, helping them to identify protection mechanisms. Lola and Anna felt that only their mum could control their impulsivity towards online shopping. Anna also reflected on developing a certain feeling of *“annoyance”* when her parents called her out for spending so much time on the phone. Ineke uses her brother to disconnect when she is fed up with online content. Conversely, teenagers rarely talk about privacy risks with their parents, focusing only on the risks of exposing themselves online or meeting new people.

7.4.2 Personal and Social Contexts Influence the Risk Experience

This theme describes how contextual elements play a role in teenagers' experiences with manipulative designs and their associated risks. Their personal context might be triggered by manipulative designs, but their social context also shapes trade-offs when they encounter these designs.

"The designs are very worthy... At the beginning I didn't want to spend on skins, but now it has become a part of my life."
(Lola)

"I just wanted to buy something 10 euros value, and my friends told me 'hey! look at this'. So to try to be a cool kid, I bought it."
(Oskar)



"It gives the exclusivity of the moment, right? It might disappear and in three years it will be exclusive."
[...]

If in two weeks I see someone else with that [skin] I say: 'I should have bought it'. [...] But you cannot do it anymore.
[...]

About exclusivity, [...] it's about thinking that you were in that moment [...] like in real life with stickers or some video consoles, things that have that exclusivity, right? It was in that moment and those that were there could get it.
(Victor)

Figure 7.2: Examples of relationships with others leading to risk in the context of limited skins on video games

Personal contexts influence emotions and shape one's sense of risk

Participants repeatedly expressed concerns about how the use of these platforms could affect their mental health and how they could see impacts on close friends and family. When participants were asked about their opinion of the design of social network platforms, two points stood out: *the design may trigger inner impulsive and addictive behaviours*, and *comparison may trigger inner insecurities*.

Participants identified the stimuli deployed on the platforms as triggers of potentially impulsive and addictive behaviors. Anna (Figure 7.3) was triggered to purchase an item impulsively when a reminder automatically added it to her basket: *"she feels like a puppet"*. Feeling bad, she wanted to exert her agency because she *"should be an adult someday and make wise budget decisions."* Similarly, Oskar felt teenagers are more susceptible to some mechanisms, like pop-ups, when purchasing. He feels *"he cannot control himself"* when windows pop up at fast-food kiosks.

Constant reminders and notifications, including unsolicited elements that pop up, seemed to affect participants, leading to self-imposed protection mechanisms.

Addiction or feeling *hooked* were common terms participants used when asked about their experiences and potential risks, expressing feelings of powerlessness.

One common risk was triggering social comparisons that might cause *mental health problems*. Women in particular worried how they could be impacted by influencers and role models. They found themselves and their friends feeling sad, comparing themselves with the content, and having negative thoughts about perceptions of their own lives and bodies. Those feelings concerned participants who explained that platforms would trigger inner insecurities. They sought to avoid seeing this type of content to reduce the likelihood of triggering these insecurities.

Context influences the interplay of resources: never money, sometimes time, and always privacy.

Participants balanced their resources in relation to their personal circumstances, which shaped the risks they took and how they perceived them. They were not willing to give money, they might give time under certain conditions, and almost always gave up privacy when they were unaware of the risks.

Teenagers' limited budget influences their conception of risk and what they can do. Participants described how spending money attaches them to video games. Spending money is a big decision, and money loss was perceived as a great risk. Participants reported "doing anything" to avoid money loss: falling for discounts and fake scarcity, seeing more ads, or spending time waiting for loot boxes containing the items they were unwilling to pay for. This sometimes led them to yield privacy or time in favor of saving money; Alex, for instance, explained how he prefers to see advertisements instead of paying.

Their perception of time as a resource influences their relationship with manipulative designs and vice versa. Participants repeatedly explained how and why they feel hooked on manipulative designs: they feel "bored" because they have free time without external regulation. Participants detailed how they got hooked in their "downtime" — on the bus, on breaks, between homework, or during the summertime. External obligations influence when they perceive "free or down" time and, therefore, the relevant risks. If they had to do something else, it was a risk; if not, it was fine. This perception of their time and lack of responsibilities contributed to having a sense of agency over what they do: they felt in control because they were managing their free time.

Other school obligations can also exert influence. Anna explained exams make her stressed and, consequently, more likely to buy clothes. They imposed some self-regulatory mechanisms to stop wasting time when they have another external obligation that outweighs it: attending school, doing homework, or spending time with family. Interestingly, Oskar explained a strategy he uses to avoid being affected by loot boxes that operate under "play-by-appointment" — meaning they can only be opened after a determined amount of time has passed. He opens them before school, attempting to match the next one after school so he doesn't risk being caught opening them in class. High school thus "imposes" on his time to open the boxes, but also what is at risk when he spends time opening them in class.

This contextual effect was also seen in instances where teenagers discovered manipulative designs. Participants mentioned having seen manipulative designs related to choice

architecture and pop-ups in the physical kiosks of fast-food restaurants, a place that is part of their leisure time with friends. However, most participants only reflected on those settings as potential triggers of behavior when the interviewer provided the examples. For participants, the presence of manipulative designs were perceived as normal, and they insisted they were used to it.

7.4.3 The (Un)conscious Experience of Harms

This theme encapsulates the different harms that participants experience when they feel manipulated in the presence of manipulative designs. Participants expressed concerns about a variety of harms related to manipulative designs and the use of platforms, including emotional distress, labour, and cognitive burdens, attentional harms, privacy harms, financial harms, as well as identity and socio-political misinformation harm. We report elicited harms from the conversations with participants. While some harms are visible to participants, others have remained unnoticed yet visible to the research team. As we build on lived experiences of manipulation, we also report some sources of felt manipulation and harm that do not necessarily relate to manipulative designs. We support this theme with Table 7.1, which maps experienced harms to documented ontologies of manipulative designs and platform affordances [172, 315].

7

Emotional distress

This type of harm includes negative emotions and psychological impacts on users — e.g., annoyance, stress, and frustration. For participants, these sneaking techniques — e.g., putting elements in the basket automatically — and attention-capture designs — like notifications — impact that emotional distress (See Figure 7.3).

As mentioned above, a common perceived harm related to triggering social comparisons that might cause emotional distress. To address these feelings, they reported avoiding seeing this type of content and disengaging with the platform by using other attention-capture design patterns — like scrolling — to reduce the trigger of those insecurities. This increased their cognitive burden, placing the effort to avoid harm on them.

Some participants also reported emotional distress — feeling bad, anxious, comparing themselves, and being mad and frustrated — when they created and posted content. As Victor explained, *“you start to see how others get a lot of likes, and you have nothing”*. Thus, when Anna realises she does not get any likes or views, she re-posts the content again because *“why did she post it for if nobody sees it?”* Some manipulative designs, like social proof and attention-capture deceptive patterns, may also negatively affect the user as a content creator and not only as a content consumer.

Labour and Cognitive Burden

This type of harm increases the effort — both cognitive and physical — on the user within the interaction. Participants reported seeing these techniques commonly when exposed to pop-

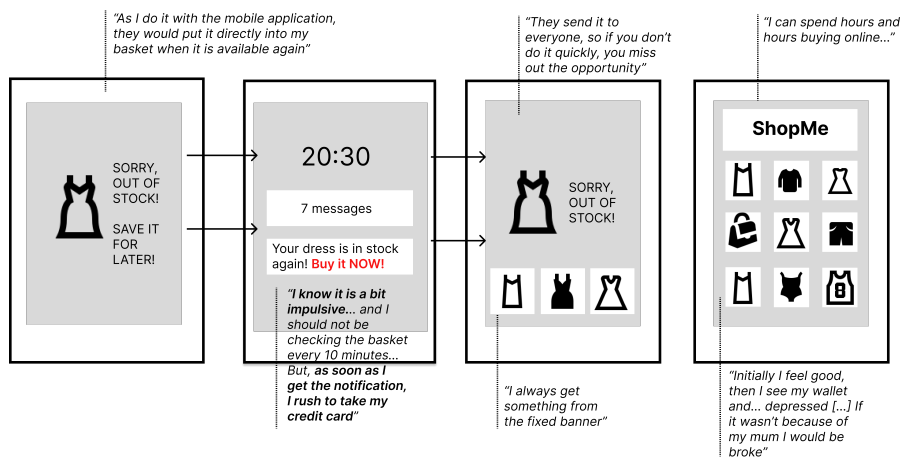


Figure 7.3: Diagram of Anna's experience with manipulative design practices triggering their impulsivity.

ups, nagging techniques, and interface interference in online interfaces. For instance, Anna used to see the consent notice from TikTok often asking to access her contacts, which tired her so she "would just accept". When we asked her about the visual manipulation of content with colours, she recalled extra barriers that would lead her to accept.

Interviewer: *Here you can see something you have told me before, that you can only accept or go to settings.*

Anna: *Exactly, that is the thing. Then, if you go to settings, I mean, I have sometimes tried to go to settings. So you go to settings, and you have another window, and another, and another, and to make it that way, you just press accept, and it's done. I mean, they make it very complicated, so you can press "accept."*

Anna's quote illustrates how she finally accepted the consent banner due to of this cognitive burden, which in turn impacts her privacy. However, privacy is not the only side effect that labour and cognitive burden can be associated with. Some participants reported how the increase in burden could have a financial impact, prompting them to buy an extra item or to feel strongly tempted to do so. Participants identified additional contexts in which they feel manipulated by techniques using cognitive burden and might have a side-effect impact: fast-food kiosks and pirated content websites. Oskar explained he would usually see interface interference techniques (including pop-ups, visual manipulation, and highlighting colours) in fast-food kiosks and sometimes find it difficult to resist the temptation, increasing his spending.

Privacy

Privacy harms imply that users make a data concession to the platform. During the first part of the interview, participants did not explicitly report privacy harms. When we exposed them to

examples of manipulative patterns, they recognized personal experiences of privacy harms. Nagging techniques on social media made participants accept permissions. Lola and Anna describe platforms as insistent, and both acknowledged not reading these notifications nor recalling what they do in such situations. Anna would accept them to not be bothered because this nagging is common and recurrent in all applications. When exposed to screenshots of friend spam – a technique asking for an acquaintance's e-mail in exchange for a life within the videogame, all participants who had played that game reported disclosing friends' e-mails and spamming them to get more lives.

When shown interface interference elements, participants mentioned a new context prompting labour or cognitive burden: pirated content websites. Participants found pop-ups that nag them that are difficult to close. Oskar explained that he sometimes receives those pop-ups while watching a movie and 'fails' in pressing the 'X' button to close the pop-up; hence, he is forcefully redirected to ads.

Attentional harms

These harms push users to spend more time by directing their attention to the platform. Participants identified attention deceptive patterns in video games, for instance, the full-screen mode that makes users lose track of time (which is a type of time fog), a notification to catch the participants' attention, grinding, and infinite scroll. Some attention capture patterns (e.g., play by appointment) were not only associated with attentional harms (see in 7.4.2 how it translates into the regulation of participants' routines) but also evoked emotional distress and annoyance.

Techniques like pay-to-win were also present. Although pay-to-win and urgency/scarcity in content — like skins — were related to financial harm, these strategies are linked to attentional harms in our participants because financial investments required them to invest more time. Lola and Victor explained buying skins as an investment because it relates to how much time they spend on the videogame: they would not buy a skin for a game they do not play.

In social networks, attentional harms reported by participants are associated with attention capture deceptive designs, mainly infinite scroll and personalisation. As a source of harm, both strategies were also mentioned in other contexts like e-commerce. Anna and Ineke explained how infinite scrolling through personalised clothes suggestions makes them invest much more time than they would initially want.

Financial harms

This refers to a financial loss for the user. Besides financial harms as a side effect of cognitive and labour burdens, participants mentioned cases of scams and deception via deceptive websites that pretend to be legitimate or dropshipping practices. These cause financial harm and emotional distress, with fear of using these platforms again. This is not a manipulative design per se but a deceptive one motivated by scammers rather than a platform. Another source of financial harm was the content influencers and video streamers promoted on platforms.

Experienced Harm	High Level Manipulative Design	Meso-level Manipulative Design
Emotional distress	Forced action Sneaking Social engineering Content/People Deception	Play by appointment* Pay to win* Confirmshaming Personalisation
Labour & cognitive burden	Forced action Interface Interference	Nagging Manipulative Visual Interference
Financial	Social engineering Forced action Content/People Deception	Urgency Scarcity Personalisation/social proof Infinite scroll*
Attention Time	Interface Interference Forced action Interface Interference Social engineering	Time fog* Grinding* Infinite scroll* Notifications Forced continuity Manipulative Visual Interference Personalisation
Privacy	Forced action	Friend Spam
Socio-political misinformation	Social engineering Forced action	Personalisation Infinite scroll*
Identity	Social engineering	Personalisation

Table 7.1: This table relates the harms reported by the participants to the high-level and meso-level manipulative design patterns described in Gray et al. [172]. Meso-level patterns marked with "*" correspond to attention deceptive design patterns, as gathered in Monge Roffarello et al. [315]. This table is not meant to be exhaustive but rather a supporting visual tool to map reported harms with associated manipulative designs.

Participants reported financial harms associated with e-commerce and video games. Social engineering patterns like scarcity claims and limited content (e.g., skins for characters and pay-to-win mechanisms) seemed a prominent source of financial harm as Lola reported (See Figure 7.2). As seen in 7.4.1, harm involves a social component with teenagers, which may contribute to its tolerance. Although participants seemed aware of social engineering techniques, they tended to accept them as an inevitable part of the game.

Socio-political misinformation and Identity harms

Participants show two additional concerns not linked to the presented contexts as part of felt manipulation. On social networks, participants reported having experienced socio-political misinformation derived from personalised content and algorithms, which create echo chambers. Lola reflected on how this information targets people akin to the content. On one occasion, she found herself conflicted when she discovered the other side of a specific political phenomenon, losing herself because she did not know who to trust. Teenagers changing their behaviour because of trends on social media is another harm that participants reported as caused by personalisation and algorithmic profiles. Ineke exemplified how this could affect teenagers' identity:

Ineke: *In some cases, [social networks] can change people's lives and make it harder or cause them problems.*

Interviewer: *What do you mean?*

Ineke: *In their way of being, their way of expressing themselves, seeing the world, everything. Maybe, someone who has a lot of confidence when talking starts seeing videos that say, 'no, you can't do this and that while talking' - because there are videos like that. [...] Then, the person feels identified with that stuff and stops doing them and stops being themselves because of the videos.*

7.5 Discussion

In this section, we describe how teenagers' experiences of manipulative design depend on contextual factors such as the social ecology of manipulative designs. We describe how the environment and ecology of teenagers might contribute to — or limit — their position of vulnerability toward manipulative designs. As one of the first accounts of the experiences of teenagers with manipulative designs, the present study provides a foundation for future researchers that includes open questions and design challenges in counteracting manipulative design experiences for teenagers.

7.5.1 The Social Ecology of Manipulative Designs

The perception of the world of teenagers is heavily influenced by their peers and families [301]: they are the immediate surroundings that help them to understand the world and shape their

experiences, including their relationships with manipulative designs.

As explained in our findings, sometimes risks from manipulative design are shared with friends and peers, incentivising teenagers to engage with manipulative designs. In this regard, in line with studies on social media use, it can be argued that manipulative designs take advantage of “network effects:” the more people engage with the platform, the more useful it is [363]. Some manipulative designs similarly rely on network effects to achieve a bigger impact; “*pay to win options*,” fake limited scarcity or social pyramids in video games do not function with only one user. This also resonates with some previous work explaining how cosmetics in videogames are not only a way of belonging and comparison with others, but also a part of the player’s identity [150, 65]. Li et al. [271] explain the hedonic aspect of in-game purchases as a consideration to include in videogames design; however, as seen in our results, when combined with manipulative designs, it might catalyse harms.

Considering our results, one may question if manipulative designs prey on teenagers’ identity or social needs more than utilitarian needs. For instance, teenagers’ experience of scarcity in video games emerges differently than scarcity cues in e-commerce due to time and social pressures, differing from adults’ experiences with scarcity that tend to focus on the product [433] and are negatively related to the hedonic dimension of user experience [439]. Addressing “network effects” and identity issues together mean that the emotional part of manipulative designs can transcend beyond a cute or sad message, like in “toying with emotions,” and rather target a teenager’s social needs and identity in the experience of the manipulative design. Consequently, the mechanisms of protection for teenagers might differ from those proposed in the literature to prevent scarcity cues and impulse buying from happening (e.g. [316, 409]). This social aspect of the experiences of teenagers with manipulative designs poses a new challenge for design countermeasures: how can design interventions include this social and identitarian aspect to protect teenagers?

Shared family experiences influence how teenagers perceive manipulative designs and their consequences. This can have both positive and negative effects on the user. The privacy literature confirms that teenagers are more aware of interpersonal privacy (i.e., disclosing information) than commercial privacy (i.e., sharing data with companies) [274], which is echoed in our results. However, if parents are equally unaware of how manipulative designs lead to commercial privacy risks, they might reinforce them. Our participants focused on how families advise and support teenagers, shaping their awareness about what they can, cannot, or should do to avoid risk online (e.g., disclosing their images online). However, accepting a banner that gives extra access to their data from other websites is not construed as a problem. This contrasts with documented experiences in adults, where effects of manipulative designs in cookie notices explains how participants reported concerns about how third parties collect data, profiling, or surveillance [46]. For our participants, this privacy issue was unlikely to be a concern unless their environment taught them that it should be, and could become more problematic than in adults given teenagers’ preference to deprioritize privacy over time or money.

Teenagers might also seek support for those manipulative designs they perceive as out of

their control. As reflected in our interviews, participants would talk about measures coming from their toolset, or their parents' support. This contrasts with other documented experiences in which participants would come up with institutional countermeasures - education, governmental intervention, or laws [287]. When teenagers sought support, parents sometimes gave advice based on negative shared experiences that ultimately might unduly impact teenagers, excluding them from the system. A family's bad experience with manipulative designs resulted in avoiding certain technologies altogether. Despite the well-intended actions of the parents, not being able to support recovery from online manipulation properly could have an impact on how teenagers develop their digital skills and a healthy relationship with the internet.

The specific trade-offs teenagers make — e.g., money, time, and privacy — will have an impact that might not be the same for other populations. Teenagers might become vulnerable to those manipulative designs that increase the cognitive burden and make a trade-off between money and privacy, or money and time, which resonates with the idea of “participatory reluctance” that Westin and Chiasson [458] pointed out: teenagers would be more likely to yield their privacy even if they do not necessarily desire it. Some direct consequences can be perceived with the new consent-or-pay model that has been implemented in Europe: if users want to use a service, consent banners offer the option of subscription to avoid the use of cookies. Given the trade-off of resources that teenagers and some other vulnerable collectives make during their online interactions, these populations may be highly impacted by such a model. Their ecologies also play a role in the effect of some manipulative designs that shape their routines, like play-by-appointment, that are specially made for them. This interplay of these trade-offs can be perceived as factors of vulnerability that do not necessarily belong to teenagers, but that seem to be present in their experiences — e.g., the lack of resources will not be only present in teenagers' ecologies. This effect of the environment as a mediator of vulnerability seems to be in line with the theorization of DiPaola and Calo [113], and it is a starting point to discuss an empirical approach to vulnerability towards manipulative designs.

Studies in design focusing on harms are growing [315, 377, 281]. A commonly suggested counter-intervention in the domain of manipulative designs is the use of friction to increase reflection and prevent users from the effects of manipulative designs. Moser et al. [316] and Bongard-Blanchy et al. [45] and Lukoff et al. [281] offer the use of friction and ‘microboundaries’ [281] to mitigate the effects of manipulative designs and help users to regain control in social media or e-commerce contexts. Indeed, Zac et al. [472] explored the use friction as countermeasure and proved to be effective in some contexts. However, in light of our results, it is necessary to consider whether this type of intervention can be adapted to the social ecology of manipulative designs and whether they would work in the particular trade-offs some populations make during their interactions with manipulative designs. To what extent is friction an appropriate solution when lack of awareness is not the main reason behind the impact of manipulative designs on specific populations?

7.5.2 Supporting Vulnerable Actors in Confronting Manipulative Designs

The effect of the social environment on manipulative designs resonates with theories of digital inequality. The effect of socio-digital inequality is well-documented [195, 197, 196] and may have an impact on experiences of manipulation. Differences in socio-economic status (SES) will impact the level of digital skills of families since higher economic resources are associated with more mental, social, and cultural resources [446], in line with the idea of socio-digital vulnerability towards manipulative designs [113]. Socio-economic conditions may, therefore, impact how teenagers deal with manipulative designs. Families with lower SES, education, quality time to give, or less tech-savvy families can misguide teenagers in their interactions with manipulative designs. Thus, the quality of the offline social network has been associated with mental health harms on teenagers when they use social media [377].

If teenagers' environments cannot supply them with the support needed to recognise or recover from the effects of manipulative designs, this can become problematic for their development, such as associations between mental health risks and the use of platforms with manipulative designs. Not being able to cope with these kinds of risks might be especially problematic for the development of teenagers' personalities. For instance, Livingstone et al. [273] documented parents being unaware of some of the risks that teenagers had experienced online. Therefore, our data hint at how digital inequalities can reinforce the vulnerability of teenagers towards manipulative designs because they attack key identity and social-related factors of their development while being dependent on others to protect themselves and recover from the impact. This dependency on the environment means that context can foster vulnerabilities towards manipulative designs for teenagers, but it also can help to prevent it. Consequently, there is room for new countermeasures to manipulation from a socio-technical point of view, both for policy implications and design interventions.

7.5.3 Limitations and Future Work

This work presents some limitations. Our sampling and time limitations did not allow us to go deeper into some of the themes, which could have shown more nuances with additional participant perspectives. Furthermore, limiting the technology contexts may have left out other relevant factors that may play a role in the experience of manipulative designs. For instance, different contexts embody different privacy trade-offs, cognitive load, and primary tasks; e.g., casual Internet browsing might entail different rationales than objective-oriented tasks.

Given the complexity of online platforms that mediate the experience of manipulation, we cannot fully disentangle the effect of specific manipulative designs from other technological affordances. For this reason, we suggest future researchers on manipulative designs to carefully reflect on the methodological implications of studying a phenomenon that cannot always be seen, like online manipulation, and to opt by taking harms-based approaches.

Our findings support a wide range of future work in the domain of manipulative designs, including the study of more diverse populations, especially those that can be considered vulnerable. Future work should seek to better comprehend the landscape of manipulative de-

signs and their effects on particular populations and more fully account for ecological complexity and sociality relating to the experience of manipulative designs.

Future areas of research may include: (i) Further exploring the role of family and friends and their relationship to coping with manipulative design, including families when exploring the solution space for interventions. (ii) Addressing the relationship between digital inequalities and manipulative designs—supporting broader goals of understanding how vulnerability impacts experiences of manipulative designs. To do so, we encourage researchers and practitioners to use participatory and bottom-up approaches that investigate "why" these harms occur, rather than maintain behaviouralist approaches that look at the effect on behaviour caused by design elements. By contextualising the interactions, researchers may elicit more vulnerability-centered counter-interventions to manipulative designs. In this paper, we interviewed teenagers about their daily interactions with manipulative designs. Aiming to disentangle their experiences with manipulative interfaces, we illustrated the impacts that their context has on teenagers' relationships and experiences with these designs. We, therefore, contribute to explaining their potential position of vulnerability towards manipulative designs, offering a new social space to fight manipulative designs that can inspire regulators and future research.

7

7.6 Chapter Takeaways

By looking at teenagers' experiences with manipulative designs, this chapter provides two important aspects of teenagers experiences that drive their vulnerability to manipulative designs. First, the social aspect — both a lack of and the dependency to it. Second, the imbalance position in which users are asked for impossible trade-offs through the user interface. Similarly, this chapter has exposed how the socio-contextual aspect can contribute to the reinforcement of inequalities and, therefore, vulnerability to these designs. Thus, while the drivers presented drive the exposure of users to these designs, the relationship with inequalities increases the impact.

Manipulative Design and Older Adults: Co-Creating Magic Machines to Understand Experiences of Online Manipulation

This chapter is based on the following peer-reviewed publication:

Lorena Sánchez Chamorro, Romain Toebosch, and Carine Lallemand. 2024. Manipulative Design and Older Adults: Co-Creating Magic Machines to Understand Experiences of Online Manipulation.

In Proceedings of the 2024 ACM Designing Interactive Systems Conference (DIS '24). Association for Computing Machinery, New York, NY, USA, 668–684. <https://doi.org/10.1145/3643834.3661513>

Abstract. Manipulative designs—i.e., dark patterns—have pervaded online interactions in most sectors from e-commerce to social media, banking, and healthcare. Understanding how individuals experience and cope with online manipulation is essential to support evolved design practices and regulatory measures. Yet studies on populations who may be more vulnerable to online manipulation are scarce. Through a series of “magic machines” workshops, we investigated the experiences of older adults (N=31) with online manipulation, their needs, and the strategies they imagine to resist manipulative practices. Our results show that participants tend to attribute manipulation to an “unknown” person and do not distinguish platforms from content. Through their machines, they expressed four primary needs to resist manipulation: knowledge, awareness, right to sanctuary, and control. Our study contributes insights into older adults’ experiences with online manipulation and brings design challenges for effective countermeasures to manipulation that address the needs of all users.

8.1 Introduction

Manipulative designs — i.e., dark patterns¹ — are design interface elements that try to steer, coerce, or manipulate users into decisions that, if well informed, they would not make [297]. Due to their potential to cause harm to users, manipulative designs are a rising concern for policymakers and scholars [341, 132, 340]. These experts bring multi-perspective interventions to tackle the problem from a design [314, 78], regulatory [181], or educational [374] perspective.

Older adults have been considered a group that embodies vulnerabilities in the online domain, as addressed by the Organization for Economic Cooperation and Development (“OECD”) report on consumer vulnerability, the European Data Protection Board guidelines [41] on manipulative design in social media, or the International Telecommunication Union (“ITU”) [72]. This in turn is likely to increase the potential harm of manipulative designs on this population. We understand that vulnerability is a multifaceted and layered concept [283], also in the domain of manipulative designs. We do not believe in labeling populations as vulnerable; however, to understand what contributes to making them vulnerable in the online domain, there is a need to explore the diversity of their experiences and needs. The rising population age and life expectancy are demanding technological interventions that adapt to older adults’ needs, as pointed out by the report on aging [72]. Investigating the experiences of older adults will allow us to deepen into their needs, so we can include them in the conversation of manipulative design. This will also help us to understand factors of vulnerability to extrapolate to other populations in the online domain.

The literature on manipulative designs from the HCI scholarship is growing [171, 339], with special attention on evaluating when an interface contains manipulative designs, and understanding users’ perceptions and behaviours when they face manipulative designs [171, 45, 296, 336, 78]. While few studies have considered age as a variable impacting the interaction with manipulative designs [45, 280, 15], there is a need of studying older adults’ experiences and including them in the conversation of manipulative designs.

Notwithstanding the narrative of older adults’ reluctance to use technology [452], this population’s needs regarding technology have been studied in the context of healthcare [66], social support [37, 36], digital banking [262], voice interfaces [54], privacy [154, 262], agency or control [24]. Hence, understanding the experiences when they feel manipulated by these technologies, might contribute to understanding their vulnerability towards manipulative designs.

Understanding manipulative designs as a problem of felt online manipulation [379, 163], we conducted a series of three magic machine workshops to understand older adults’ experience of online manipulation. We collaborated with three social organisations that facilitated the recruitment of the participants. As speculative methods, magic machine workshops are

¹The research community is studying this phenomenon using a variety of labels, including deceptive design, sludges, manipulative designs, and most dominantly, “dark patterns.” We hereby use the term “manipulative designs” to describe this phenomenon, embracing the critics towards the term “dark patterns” and following the ACM recommendation on diversity and inclusion [143]

meant to give voice to participants [10]. By asking participants (N=31) to create a "magical artefact" to be protected from online manipulation, we elicited their relationships with and needs towards online manipulation.

This paper provides several contributions to the HCI and design community. To the best of our understanding, we document the first study on older adults from the perspective of manipulative designs and online manipulation. We provide empirical insights into the understudied experiences of older adults with felt online manipulation by documenting their needs and lived experiences of felt manipulation, as well as their relationship with the ecosystem of manipulation. We discuss how to include older adults' needs in the conversation of manipulative design and provide some design challenges to create countermeasures. Our findings represent a starting point in understanding the ecologies of older adults in the realm of manipulative designs and open the discussion of how to better protect these populations. By expanding our understanding of the role of older adults' ecologies into the experience of online manipulation, we hint towards potential factors of online vulnerability. We expect to inspire future research on vulnerabilities within the community on manipulative and deceptive design.

8.2 Related Work

This section frames the problem of manipulative designs, with attention to the concept of vulnerability. First, we explain why understanding the experience of online vulnerability in the realm of manipulative designs is important. Second, we illustrate how the ecologies of older adults play a role in their interaction with technologies. Third, we provide an overview of how speculative design methods have been used to elicit user needs in older adults.

8.2.1 The Experience of Manipulative Design Practices

Manipulative designs are complex, pervasive, ubiquitous, and use subtle mechanisms [379], whose study benefits from a harm-based approach to understanding the ecologies of users' vulnerability. Despite manipulative designs can lead to harm [139, 181, 341], users present difficulties in identifying such design strategies [308, 45, 287, 15]. Maier and Harr [287] explained how, when shown examples of manipulative designs, participants evaluated their severity in relation to their potential harm. However, as explained by Bongard-Blanchy et al. [45], even when users are exposed to manipulative designs and can identify them, they are not sure about the harms these designs can cause them — like the risk of manipulation. With this rationale, Gray, Chen, et al. [163] used the proxy of felt manipulation on users to extend the analysis of manipulative design effects. Hence, taking a phenomenological approach to looking at when users felt manipulated by the designs might help unveil their experiences.

Inspired by science and technology studies terms [424], we understand manipulative designs as relational; users only perceive them when an interaction turns into a negative outcome. Gray et al. [169] anticipated this idea with Norman's gulfs since manipulative designs

are perceived as a mismatch between users' expectations from the interaction and the interaction itself. Gray, Chen, et al. [163] also explored the idea of "temporality of manipulative designs." This phenomenological approach aligns with our goal of understanding the experience of older adults via the proxy of online manipulation.

In the last few years, the study of manipulative designs has increased [339, 171]. A wide range of literature has focused on how these mechanisms impact users' behaviour in different domains — such as privacy, digital services, streaming platforms, and social media. While some studies have explored how users' characteristics influence the experience of manipulative designs like education [45, 280], and age [45, 15, 447, 382].

A minimal set of literature has explored the effect of age on users' interactions with manipulative designs. Bongard-Blanchy et al. [45] reported, via a survey study, that older generations had more difficulty identifying manipulative designs than younger ones². On the contrary, using an experimental setting, van Nimwegen and de Wit [447] found a negative association between age and falling for the specific manipulative designs — i.e., sneaking extra products or cost into the users' basket and use of emotional manipulation. Their experiment combined both types of manipulative designs and showed how older users were less likely to fall for those. While Avolicino et al. [15] explored vulnerability towards some manipulative techniques and age, their analysis was limited to two age categories: under and over 35 years old. They find the latter more vulnerable for being less acquainted with the service and mainly experiencing anger, sadness, and disgust. Given the minimal and non-conclusive findings about age, we aim to account for older adults' manipulation experiences to understand their unique ecologies that might contribute to vulnerability. These ecologies are different nested structures surrounding a person [57], which might impact their experience. Understanding the experience in the context of manipulation is the first step to understanding their ecologies and informing design interventions that also include older adults' needs.

8.2.2 Factors of Vulnerability in Older Adults

The concept of vulnerability is a multifaceted construct that has been discussed from theoretical and legal perspectives [283, 136, 289, 340]. Vulnerability translates into being more prone to a position of susceptibility towards a threat, being less likely to recover from its impact given a position of power imbalances [283, 288]. We do not intend to deepen the debate of vulnerability, but for the sake of this study, we understand vulnerability as layered [283]. Vulnerability is not an absolute construct but situated; some users are more vulnerable than others given intersectional conditions — e.g., race, age, or gender. We take this stand of vulnerability in the domain of manipulative designs: vulnerability towards manipulative designs is layered and does not exclusively depend on one label. In the realm of manipulative designs, the extent to which users can be more vulnerable given their personal conditions — such as age, gender, health, and socio-economic background [472, 113] — is raising more attention, creating the need of a deeper understanding of users and their potential conditions

²This study counted on the baby boomer generation and above, including participants above 56 years old in their sampling.

of vulnerability.

Policymakers and researchers have considered older populations as potentially vulnerable to manipulative designs [41, 340]; yet the reality about these populations in the context of manipulative designs is still uncertain. Understanding older adults' experiences with online manipulation will give an insight into their ecologies and what makes their situation unique when facing online manipulation, contributing to debunking assumptions and initiating design interventions that consider their needs.

Van Deursen and Helsper [449] explained how older adults over 75 feel “too old” (p.182) for using the internet, showing there are differences between gender, education, and age. Additionally, Barros Pena [24] explained older adults reduce their sense of control and competency over technology over time, which motivates this disengagement. In older adults, in general, self-efficacy is an extended factor documented in the literature [317, 449, 357]. In low-income adults, self-efficacy, distrust, perceived risks, or social support also affect the use of technologies [75].

The pervasiveness of technologies creates more opportunities for older adults to engage in digital activities [72]. As reported by Van Deursen and Helsper [449], leisure and recreation online activities are predominant in male older adults, while social online activities are more common in females. “Cognitive-knowledge enhancing activities” [449](p.182) online, like news consumption, are correlated with highly educated older adults. Several studies have presented the needs of older adults within their online interactions, like privacy and information sharing [154], or social connectivity [72] with the use of social platforms. As Barros Pena et al. [24] pointed out, the study of older adults' experiences and needs with technology has to be understood beyond usability and accessibility. Thus, the predominance of manipulative designs in platforms related to older adults' needs — like social media, video platforms, or newspapers [180, 309, 308, 401, 171] — requires understanding older adults needs in relation to these manipulative designs.

8.2.3 Why to Use Speculation to Talk About Manipulative Designs with Older Adults?

The “Magic-machines” workshop is a speculative method that builds on performative techniques, aiming to inverse roles between facilitator and participant, and allows participants to take the lead [10, 11]. The exploration of refugees [7], LGBTI community [157], or older adults' needs [40, 277, 452], are some of the examples that used this method in different contexts such as immigration programs [7], meditation [104], period-positive technologies [69], urban planning [40], data-driven technologies [18], parent-children communication [318], the practice of care [277], hate crime reporting [157], banking [452] sensors [332], music [268], and mental health technologies [190]. From these documented works, we learned that this method is suitable for populations whose voices are less recognised, to talk about technological needs without the hurdles of speaking about technology.

As the subtle techniques that manipulative designs use are often not perceived by users,

Andersen's concept of magic machines [10, 11] seemed appropriate for our goal in two ways. First, the "magical" approach of the method would allow older adults to talk about their experiences via the proxy of felt manipulation and the consequent harms they experienced. Second, the "performance of the method," namely explaining in an improvised way how machines work and how they would use it against manipulative designs, would allow participants to talk about their concerns about manipulative designs in ways they did not even know; it is the moment in which the internal monologue of participants goes on stage, and it is orally expressed.

To the best of our knowledge, little research in the field has looked into older adults and their experiences with online manipulation from the perspective of manipulative designs. Yet it is fundamental to understand older adults' experiences and ecologies with online manipulation to comprehend their vulnerability towards manipulative design. Hence, this study addresses the following research questions: **(RQ1) How do older adults experience online manipulation in their everyday life? (RQ2) What are older adults' needs when it comes to addressing manipulation online?**

8.3 Methodology

Inspired by the work of Andersen and Wakkary [11], we conducted a series of "magic machines" workshops with older adults on the topic of online manipulation. As a speculative design method, magic machines workshops are used to empower users by trying to shift the leadership of the workshop to the users. They help situate users in "magical" contexts, so they do not need to talk about technology attached to existing artefacts, avoiding their hurdles but allowing the user to explore without limits while expressing their needs. For this reason and as evidenced in prior work, magic machines workshops are a suitable method to work with users who might be in situations of vulnerability, like refugees [7], older adults [40, 277, 452], members of the LGTBI+ community [157].

8.3.1 Participants

We involved 31 older adults in 3 magic machines workshop sessions. Our participants were aged between 61 to 96 years old ($M = 75.9$ years old, $SD = 8.61$ years old). Except for one, all participants in our sample were retired. While this study does not intend to be representative nor exhaustive in the sampling, Table 8.1 provides some socio-demographic information that helps to understand participants' backgrounds and the overall composition of the sample. It includes basic demographic information such as age and gender, along with their reported education, socio-economic conditions, and profession before retirement.

To facilitate participation, we used snowball sampling to contact social organisations whose primary mission is to work on older adults' advocacy. Three organisations expressed interest in the study — two local NGOs and a daycare centre — and promoted the activity among their usual members. The organisations disseminated the workshop information via posters and

word-of-mouth³. On the day of the workshop, the participants interested in the activity would join us. The inclusion criteria required participants to have a mobile phone with internet and to use internet to make or receive communications.

- The first group comprised members (n=11) of a neighbourhood association from a working-class neighbourhood in Madrid. They were older women, most of them widowed. The association arranges activities for them to talk about various issues and create a supporting network. They meet every Monday to participate in different activities, from discussions to more creative activities — e.g. painting. Most of these women declared using the internet to communicate with family and to seek information online (e.g., searching for recipes). All of them embraced the activity similarly to any other weekly activity, with curiosity, but also with the purpose of having fun as a primary aspect.
- The second group involved members (n=9) of another neighbourhood association from a different working-class neighbourhood in Madrid (6 identified as female, and three as male). There was some heterogeneity in the group composition. Most of the participants were active members of the association, contributing actively to the local development of the neighbourhood, and decided to join the activity. Some participants were attracted by the activity itself as an opportunity to know better the organisation's work or were invited by other association members. There was only one participant who did not know anybody in the group, and who expressed having difficulties embracing the creative aspect of the activity. This group reported using the internet daily to be informed (e.g. reading newspapers or searching for recipes), but also to disseminate information they find useful or interesting among family, friends, and neighbours. They spend some leisure time on the internet and, sometimes, play. Administrative issues relating to banks or insurance companies and managing e-mail were also among their usual activities. This group also pointed out how “their generation preferred to talk” over texting, explaining how conversations over texting lead to misunderstandings.
- The last group was composed of members (n=11) of a day-care leisure centre from Salamanca (10 identified as female and one as male). They are regularly involved in activities within the day-care, where they have full autonomy and spend their leisure time, for instance, playing bingo regularly. There was a lot of diversity in the usage these participants made of the internet. One participant reported using it for mobile video games, social media, and information searches in general. Most participants reported using it for information searches —celebrity-related content or cuisine. Some participants also reported using e-mail. One participant reported only using it for communications via WhatsApp. Two participants did not see themselves ready to build a machine but fed the conversations and actively contributed to the discussion.

³An example of a recruitment poster provided to the organisations can be found in Appendix

The study was approved by the University of Luxembourg Ethical Review Panel, and informed consent was collected from participants, who were fairly compensated for their time.

8.3.2 Protocol

Two of the co-authors facilitated the workshops. One facilitator was a local and a native speaker of the participants' language: they ran the session and prompted participants to explain their machines. The second facilitator, less proficient in Spanish, helped with logistics, reflexive note-taking, and added questions to the participants in addition to the main facilitator's ones. Considering prior work showing how some personal conditions affect how older adults perceive themselves as less tech-savvy, especially older women [449], we decided to ask them to work in pairs so they felt less intimidated by the task. A total of fifteen magic machines were thus created across the three sessions. Each workshop entailed five parts. The full protocol description is provided as Supplementary Material.

1. Introduction and warm-up. The two facilitators introduced themselves and the purpose of the study. The participants were provided with workshop instructions to set the tone: *“(i) participants are in control, the researcher is just a facilitator, (ii) there is nothing right or wrong, everything works, (iii) everything you want works, there are no limits”*. As a warm-up activity, the participants were invited to introduce themselves briefly and possibly share with the group a superpower they would like to have.
2. Setting the stage: experiences with online manipulation. The facilitator invited the participants to explain the contexts in which they used the Internet (e.g. When do you use the internet? How? On which device? What do you normally do?). To recall moments in which they were manipulated, participants were then prompted with some manipulative design harms related to their use of technology documented in prior work with older adults ([449]): autonomy, privacy, and financial loss, as explained in Gunawan et al. [181].

We used the following prompt:

“When was the last time that you were on internet/or mobile phone and you did something you did not want to? Or did not expect? Or perhaps even regretted? It could be one specific thing or something that happens more frequently to you.

Maybe you gave your privacy unintentionally? Subscribing for an unexpected service or newsletter... or someone called you trying to sell something, when you do not even know why they have your telephone number and data?

Even in some occasions you could spend more money than you expected, paying unexpected extra fees or buying something in a rush you didn't really need. Oh! And how many times you have spent more time than you expected on a website or app? Do you remember some of those situations?

After the prompt was presented, participants were given a paper with a sentence to

ID	Age	Gender	Job before retirement	Education	SES
1	66	Female	Cleaning staff	Secondary education (A-levels)	7
2	73	Female	Homemaker	Primary education	6
3	77	Female	Homemaker	Primary education	5
4	61	Female	High-school teacher	Bachelor's degree	7
5	75	Female	Seamstress	Primary education	6
6	61	Female	Teacher	Bachelor's degree	8
7	85	Female	Homemaker and cook	Primary education	4
8	-	Female	Seamstress	Primary education	-
9	72	Female	Elderly caregiver	Secondary education	5
10	68	Female	Early Childhood Education Assistant	Primary education	5
11	96	Female	Homemaker	Primary education	-
12	75	Male	Teacher	Bachelor's degree	6
13	87	Female	Teacher/Nun	Phd Studies	6
14	77	Female	Tourism Sector	-	6
15	69	Female	Doctor	Bachelor's degree	6
16	94	Male	Advertisement, teaching and others	Bachelor's degree	7
17	66	Female	Cleaning staff	Primary education	3
18	74	Male	Driver	Secondary education	7
19	87	Female	Music Teacher/Nun	Secondary education	-
20	73	Female	Statistics Technician	Bachelor's degree	7/8
21	74	Female	Entrepreneur	-	8
22	72	Female	Homemaker	-	6
23	70	Female	Leather worker	Secondary education	Low
24	70	Male	Self-employed	Primary education	7
25	70	Female	Homemaker	-	7
26	75	Female	Homemaker	Primary education	1
27	81	Female	Bag maker	Secondary education (A-levels)	5
28	75	Female	Shoemaker	Primary education	5
29	67	Female	Attendant	Primary education	6
30	91	Female	Homemaker	Primary education	5
31	79	Female	Clinical Assistant	Secondary education	6

Table 8.1: Description of self-reported demographic data. The table is divided by group. More detailed information can be found in the Appendix. For self-reported education, participants were asked to mention which was the highest level of education they achieved. For self-reported socio-economic status, we asked participants to position themselves on the socio-economic scale, with 1 being the lowest and 10 being the highest. One participant reported "Low" and another one "7-8"

complete individually (on paper or orally): *“The last time I did something I did not want to or did not initially intend to do on the internet was... because...”*

3. Building a magic machine. Participants are asked to create a magic machine with ‘magic materials’ in pairs. The goal of the machine is to help them to prevent the situation described in the sentence completion exercise.
4. Pitch and questions. After creating the machines, the facilitators asked the participants to present their machines to the group and explain how they worked to address their problems. Participants are invited to explain their machines, give them a name, and answer the questions from the facilitators and other participants. Participants were not previously told about it yet as Andersen explains [10], this is a crucial part of the method that allows participants to move from their internal monologue and personal decisions to an open conversation.
5. Debriefing. The facilitators thanked the participants for their collaboration and explained their research on online manipulation and the importance of older adults’ needs in addressing this problem.



Figure 8.1: Main steps of the study protocol

8.3.3 Data Analysis

The facilitators took notes from the participants’ descriptions of their internet and technology use to inform about their profile (Table 8.1). The presentations of the machines by the participants were recorded and transcribed into Spanish by the first author, and translated into English before analysis. In case of ambiguity, the original language was used to grasp the meaning of participants’ verbatim quotes.

Transcripts were analysed using Braun and Clarke’s thematic analysis approach [52, 50]. The transcriptions were coded using different strategies to enrich the analysis [378]. *In vivo* coding takes literal excerpts from participants, which becomes fundamental to give them a voice in their own terms — e.g., *“I have been scammed by e-mail, with a gift.”* *Process coding* encapsulates processes and actions within the studied phenomenon. For the context of a magic machine workshop, this type of coding is helpful to unveil user needs via the affordances of

the machines; in other words, by understanding what the machine can do for the participants — e.g., *hiding, blocking, filtering*. *Descriptive* coding summarises the meaning of a specific phenomenon shared by the participants — e.g., *They don't want to see bad messages*. Lastly, *versus* coding looks at the data from the perspective of a dichotomy or concept opposition — e.g., *the services I use VS the services I don't use*, which enriched the analysis by providing specific conflicts in the experiences of older adults.

All authors read and took notes about the transcripts to familiarise themselves with the data. Upon discussion, we elicited initial key constructs (e.g., context of use, social aspects, personal) to drive the analysis. Around those constructs, the two authors who facilitated the workshops conducted independent rounds of analysis to define themes. In the first round, the first author analysed transcriptions from the three sessions, while the second analysed one session. Using conceptual maps created by MAXQDA, both authors contrasted their analysis to develop the first initial themes. During the second round, the first author re-analysed the data set to check its suitability with the initial themes. The second author analysed a different session from the dataset. Via iterative discussions among all the authors, the initial themes were refined into two final themes and their corresponding subthemes, which address our research questions. We support our analysis with quotes from participants and explanations of their machines. A brief description of the machines can be found in the Appendix.

8.3.4 Positionality and Ethical Considerations

As co-authors engaged in this collaboration, we harness both commonalities and differences in our research and disciplinary expertise, as well as in our positionality in relation to the work conducted. Beyond self-reflexivity, we engaged as a team in discussions around positionality, following Kohl and McCutcheon's concept of "kitchen table reflexivity" [238]. Acknowledging the situated nature of knowledge, we hereby offer some insights into these reflections around our identities in relation to our research participants.

Studying manipulative designs and their potentially harmful effects on populations likely to be vulnerable online has been a personal interest and a strong drive for the first author of this paper, who shares a socio-economical and cultural background with the participants of this study. Conducting this study in Spain, in neighbourhoods similar to where they grew up, resonated further with them, who also had personal ties with one of the organisations. Although being an outsider to the age group, the first author recognised in participants some of the experiences lived by their immediate relatives. As highlighted in prior work [238], sharing the same language as the participants is essential here to build rapport and grasp the meaning of the concepts imagined by our participants. One such example is the machine "El Cortijo," which refers to a specific Spanish type of traditional rural dwelling.

The second author is a design researcher. They do not share the socio-economic or cultural background of the participants. In their work, however, they challenge technological paternalism. The magic machine method, therefore, resonated to put the participants in the foreground. Less personally involved with the participant group, they offered insightful contrasting perspectives and a more analytical approach to the interpretation of the findings.

The last author has not been in contact with the participants. While she does not share their cultural background, she comes from a similar socioeconomic background. As a professor, she is an example of intra-generational social mobility, yet her close relatives are likely vulnerable online and often refer to her as a source of information against manipulative designs. Her main research expertise lies in design methods and how they can be instrumental in understanding participants' lived experiences.

As authors, none of us belong to the participants' age group. We focused on older adults and "set the users on stage" because we valued their voices in this discussion. We believe participants could benefit from this study since we aim to include their perspectives in the conversation of manipulative designs and how to design better countermeasures. We acknowledge the critique about "vulnerability" as a stigmatising label for older adults [283]. They are not a homogeneous group that a number can define [29], but we understand they share ecologies as they enter a new stage of life. We have tried to look through the data cautiously and consciously to prevent ageist approaches. Our differences might nevertheless have impacted our attempt to build rapport and fulfill participants' needs. Anecdotally, some participants commented on the absence of sugar-free candies during the workshop, something that none of us had considered.

We expected to find an intersection of potential vulnerabilities that play a role in how participants perceive discussing online manipulation as a sensitive topic; therefore, we adopted "care" [249, 248] as a fundamental value in this study. This was translated into considerations during the recruitment, facilitation of the workshop, and debriefing. We relied on NGOs with extensive experience working with these participants to support their recruitment and ensure that their expectations were met. During the session, we explained the consent form and reminded the participants of their rights, emphasising the right to withdraw. We presented ourselves as researchers but also as locals — in the case of the first author — with a huge commitment to include older adults' needs in technology design. We were careful and active listeners when it came to understanding their experiences of manipulation, especially those that differed from our mental models. The method we chose to engage with this topic also goes beyond "practical ethical duties" by taking care of how participants' work is presented, valued, and documented [11]. Last, in our debriefing, we discussed the implications of manipulative designs in participants' lives and their value in our study and research in general. This was important as some participants doubted the value of "making magic machines".

8.4 Findings

8.4.1 (RQ1) Imaginaries of Online Manipulation

The concept of online manipulation that participants presented transcended platforms and focused mainly on the content they interacted with and the people behind it since manipulation was understood very abstractly. These imaginaries of manipulation reflect participants' perceived agency towards online manipulation and technology in general. They expressed a

feeling of powerlessness in which technological development is described as too fast for them, which makes them feel like outsiders from the group of “those who know how to interact with technology.”

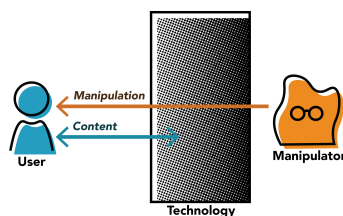


Figure 8.2: A shared conception of online manipulation amongst our participants. Participants see the manipulator as an unknown, abstract entity that manipulates users through the technology. The technology is nontransparent, and there is no distinction between content and platform

There is someone unknown behind manipulation

Participants showed two abstract ways of interpreting how they experience online manipulation, which we synthesise in Fig. 8.2. For them, *there is no systems, only people, and there is no platform, only content.*

Participants expressed their idea of manipulation as an abstract concept, hinting at an unknown “someone” behind it. In the creation of the machines, we saw a trend of blocking and preventing communication to be unreachable (See 8.4.2). However, they are unsure of *who* or *what* is trying to reach them. Participants tend to see people — either identifiable persons or abstract entities — as perpetrators of manipulation.

This abstract level of thinking is visible in how they refer to perpetrators of manipulation. Several participants had suffered scams, via e-mail, WhatsApp, or phone calls. Some referred to specific persons and even reported their cases to the police. Yet, most participants perceived the scams as executed by abstract entities. For instance, P24 talked about undesirable messages he receives from the National Post Service to pick up a parcel. He felt annoyed, and he said he knows that it is “only spam” because he did not order anything, so he does not understand why unsolicited service providers send him notifications. Yet he mentioned to have understood similar messages from Amazon, where he makes online purchases. In our view, what P24 did not seem to realise was that the messages he describes are not spam but phishing scams, and are not sent by the National post or other companies whose names are displayed in the messages. P6 alluded to manipulation through information and news, regardless of the political ideology, as something omnipresent. When discussing cookie consent banners or other examples of platform affordances and design strategies, participants used the same level of abstraction.

In addition to that *abstract manipulator*, the interface also remains abstract; there is no platform, only content. Participants’ perception of online manipulation is about the content



Figure 8.3: Examples of machines created by the participants. From left to right: “El Cortijo”, “Time Machine”, “Parabolic Antenna”, and “Stop-calls Windscreen”

they consume on the news, social media, or via personal communications. Filtering or blocking content was the purpose of twelve out of fifteen machines they created. For instance, “Magic Wall,” “Online Selector,” or “El Cortijo” would filter the content they watch on the tablet or mobile phone in real-time. “The Blind” and “Machine Not to Hear What You Don’t Want” can also be used for content coming from personal communications. “Parabolic Antenna,” “Stop-calls Windscreen,” or “Tunnel of Steel” address directly personal communications via phone. “Spam Eraser” is used for e-mails, “Magic Between Two” is used for personal communications, and any content on the interface, like “Sponge,” could be used for deleting cookie banners. A description of the machines can be found in the Appendix.

8

I don’t belong here; what should I do?

This theme gathers participants’ perceptions of themselves when it comes to facing online manipulation attempts, and technology in general. We observed an ambivalent feeling of agency and belonging across participants. They expressed that technology goes too fast, and framed their discussions by considering themselves as outsiders from a group of “people who know about technology.” Tied to the feeling of not belonging and powerlessness was a sense of responsibility for their mistakes when facing manipulation. Participants dealt with these feelings through two contrasting positions: fighting or embracing that separation.

Fighting the separation. Participants felt capable of protecting themselves from manipulation to some extent. P3 explained how she had implemented a way to avoid undesired land-line phone calls, and P24 expressed her frustrations with erasing spam messages manually. Yet, some of them wanted a “boost” to fight the separation and feel included. We purposely use the word “boost” because participants would not see themselves as “incapable” but want their needs for external support to be acknowledged. The “Flying Balloon” machine displays a message to politicians to advocate for *“inclusive access to technologies for all older adults.”* This idea of boost was also very present in the conversation between participants about the daily burden of cookie consent notices. If P20 and P15 complained already deal with this problem daily, they require a boost to help them with this tedious task.

Participants’ quest for knowledge also reflects this need for a boost. During the conversa-

tions and the creation of the machines, participants expressed the need to know more about how technology works. They feel technology goes too fast (*“things have changed overnight”*) and they find it hard to catch up. P18 recalls how things changed quickly during the COVID-19 pandemic; appointments had to be made by phone, and phone trees on mobile would confuse them because there are no physical buttons to press on tactile phones.

Embracing the separation. On the other side of the spectrum, some participants embraced the separation by blocking, hiding, and not wanting to engage with the technology. They would rather not seek adaptations or boosts to cope with manipulation and do not feel the need to be part of the group of people who know how to use technology. We observed this trend in their desires to hide from manipulation and technology, associated with their feeling of powerlessness and frustration. For example, the “Machine To Avoid Hearing What I don’t Want to” is meant to block without concession anything that might come.

(P29) Yes, for everything I don’t want to hear I have the headset. I put it on, and even if it is my neighbour, if I am not interested in what she is talking about, why would I bother? No, no!

These ways of dealing with feelings of not belonging and being responsible for one’s mistakes were present in the conceptualisation of the machines. In the next section, we describe participants’ strategies to cope with manipulation.

8.4.2 (RQ2) Strategies Imagined to Cope with Online Manipulation

We reflect on participants’ machines from the lens of the strategies identified to cope with online manipulation. We conceptualised these strategies based on how the users interact with the manipulation ecosystem and how they perceive it. Hence, we observed patterns around the relationship between participants and online manipulation through the machine — i.e., what they wanted to change in their relationship with technology or how they attempted to protect themselves with the machine. Our aim was not to systematically classify machines per strategy but to elicit strategies from the machines’ creation, envisioned use, and user interactions. Some machines can thus exemplify multiple strategies. The four main strategies identified (Figure 8.4) are insightful insofar as they embody and unveil underlying needs for *knowledge* (knowing how technology or manipulation works), *awareness* (being aware of when manipulation takes place or the countermeasure is intervening), *right to sanctuary* (the right not to be reached by technology), *control* (being in control of how one engages with the technology and the manipulation), and *the need to be heard* (by those that have the power to change things). While acknowledging a nuanced view of participants’ needs in relation to the strategies they imagined, we present predominant needs per strategy for the sake of clarity.

I want to learn

One strategy adopted by our participants acts at the individual level. To fight the separation, they feel the need to know more and decipher the system, technology, and people behind

Machine Name	Brief Description
Time Machine — <i>Máquina del Tiempo</i>	Pressing one of the three buttons would allow you to travel in time, go to the moon, or go to hell to escape. When ready to stop hiding, you can press the button again.
Parabolic Antenna — <i>Antena Parabólica</i>	This machine is made with isolation material that blocks undesired communications. Communication is only allowed with a passcode. It is portable and can be placed wherever communication may happen.
Filter for Greasy People — <i>Filtro Para Grasosos</i>	Every time the user receives unwanted communication, it gets trapped in a filter so they cannot reach anyone else. Once the filter is full of “unwanted communications,” users would replace the filter. This can be placed anywhere.
The Dwelling — <i>El Cortijo</i>	Placed somewhere near the source of communication to filter and block unwanted communications — e.g., websites or devices. The device would notify the users by moving the feathers.
Stop-calls Wind-screen — <i>Parabrisas Parallamadas</i>	This mobile device can be transported anywhere the user wants. Unwanted communications — both incoming and outgoing — would be blocked. When unwanted calls come in, the device signals it with movement.
Online Selector — <i>Selector Online</i>	This machine comprised two types of artefacts: earplugs and antennas. The device detects and blocks unwanted content by detecting the user’s brainwaves — when they perceive the content is unwanted.
Magic Wall — <i>Pared Mágica</i>	This is a device composed of a magic audio headset and visual filter to block unwanted content. Audio headsets are individual, but the visual filter can be shared using a personalized device to connect to the filter.
Magic Between Two — <i>Magia Entre Dos</i>	This machine draws inspiration from children’s string telephone kids using cups or tin cans connected with a string as a simple communication system between two people. Unwanted communication can be blocked by covering the cup.
Sponge — <i>Espanja</i>	This machine filters out any unwanted interaction or communication. Once filtered by the sponge, it can never return.
Tablegic — <i>Tablegic</i>	This machine acts like a blackboard and can be connected to any device. The user can ask Tablegic how to perform activities that they do not know how to do on their device, and it will remember when actions have been performed in a certain way.
Steel Tunnel — <i>Túnel de Acero</i>	A mobile device can be inserted into the machine, connected to a computer running a special application. It automatically blocks unwanted communications and transfers the information to the mobile. Once users remove the device from the machine, they will never receive these unwanted communications again.
The Blind — <i>La Ciega</i>	This machine is a mask that completely covers the user. If unwanted communication is received on their mobile phone, the user wears the mask to block it. Once the communication is gone, the user removes the mask.
Spam Eraser — <i>Borador de Spam</i>	This machine automatically removes spam e-mails from their inbox. Users would download and install that machine as an application.
Flying Balloon — <i>Globo Volador</i>	This machine displays a message, “For technology to reach all older adults,” to politicians and experts in the form of a permanent balloon in the sky.
Machine To Avoid Hearing What I Don’t Want To — <i>Máquina Para No Oír lo Que No Interesa</i>	This machine is comprised of audio earbuds capable of blocking all unwanted communications and messages.

Table 8.2: This table presents the machine names and briefly describes how they function. We have provided the original Spanish names of the machines given by the participants, which, in some cases, contain nuanced idioms and meanings unique to Spanish.

the manipulation. They would in turn move from an abstract representation of the mechanisms behind manipulation to a more actionable understanding. The “Tablegic” machine for instance would teach participants whenever something they face things they do not understand, including an intent of manipulation — e.g., by explaining cookie banner content. P18 would feel hopeless to let the machines do everything for them:

(P18) The machine may know a lot, but we need to learn and not let the machine give us everything done, because if we are given everything... it's over.

Similarly, P1’s “Time Machine” would allow her to know what had happened when the manipulation or deception occurred so she could learn for the future. In “Stop-calls Windscreen,” for instance, the participants wanted to also prevent themselves from making mistakes that led them to ill-intended situations. Users’ interest in learning and changing their behaviours is key in this strategy, which acts on them rather than the technology itself.

We, therefore, see the *need to know* in two ways: knowing how to interact with the technology and how the manipulation system works. This would allow participants to gain control of the manipulation and avoid it.

I don't want to be alone

In this strategy, participants act on their ecology rather than the system, the manipulator, or the technology. Participants strive to belong to the group for which they feel like outsiders. They would not be alone anymore when facing manipulation. P28 showcased the “Flying Balloon” that made politicians and technologists aware of older adults’ needs. During the presentation of this machine, some participants expressed how “for older adults [these technologies] are absolutely uphill” (P28) and that they “need someone next to me [to manage them]” (26). About Spam Eraser, P26 expressed that erasing spam from her e-mail would be done by her children because, in her words, she “is zero [at doing it].”

(P28) I have made a balloon so all technologies reach all older adults, and those would reach us more easily, and we'd not have so many many barriers everywhere. Everywhere, above all the bank, above all, a looot of things we don't understand. And you know who solved it for us? Our children and grandchildren.

(P26) And if you don't have children or grandchildren...

Although only one magic machine directly embodies this strategy, the idea of not catching up and seeking assistance was very present during the conversations with participants (See section 8.4.1), giving rise to the *need to be heard*.

I want to remove the manipulation

In this strategy, instead of changing the user, the ecology, or the technology, participants add a layer between the interface and themselves to get rid of the source of the manipulation. This strategy materialises in two ways: by removing manipulation and by blocking manipulation.

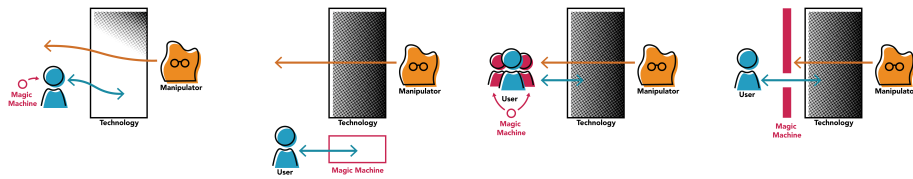


Figure 8.4: Figure 4. Visual representation of the main strategies imagined to cope with online manipulation. From left to right: “I want to learn,” “I want to replace technology,” “I don’t want to be alone,” “I want to remove manipulation”

Participants remove the manipulation and associated content directly so it cannot reach anybody else. “Sponge,” “Tabletic,” or “Filter for Greasy People” illustrate this strategy. P20 and P15 explained how to use the Sponge to remove cookie consent banners.

(P20) *Yes, I mean, when I see something that starts saying “accept, or whatever.” When I see something [with this machine], I remove it. But I remove it forever; it has to be something permanent in which something I have already removed cannot come back to me.*

(Facilitator) *What do you mean “accept, or whatever?”*

(P20) *I mean, I don’t know, every time you go online to see whatever, it nags you; I don’t know, whatever, “do you accept? Yes, no?” and you have to say “yes, no” because otherwise you don’t continue. Then, you accept. This is such a drag...*

(P15) *Yes, the thing with the cookies is killing*

(P20) *In the newspaper [the pop-ups] accumulate, accumulate and accumulate and, in the end, is impossible to read three lines without something popping-up. Then, you remove it forever [with the machine].*

Blocking manipulation was present in twelve out of fifteen machines, with some nuances. Participants expressed their need not to be reached by any source of manipulation. This strategy resonates with the *right to sanctuary* described by Zuboff [480] as the safe place users deserve in which they cannot be reached by technology. Discussing “Time Machine,” P4 explained how much she loved the idea of being invisible with the machine, “so nobody could know any of my data, nobody.” The creators of “Time Machine” not only enabled a program to travel in time but also to go to the moon or hell, as an illustration of how important it is for participants to be unreachable. The names of several of the machines directly evoked this need (Figure 8.2): “Machine not to hear what I am not interested in,” “Stop-calls Windscreen,” “The Blind,” and “El Cortijo” which refers to a specific type of enclosed rural dwelling.

(P9) *Because it is enclosed.*

(P6) *Since we started to make, she said, ‘look at it, El Cortijo!’ It is El Cortijo for the shape, which is like...*

(P9) *Like very private, a ‘cortijo’ is something very private [...]*

(P6) *That's is! The idea of the machine is privacy, right? Then we have taken these two sponges that are rough and that would filter with this foam. That's the name, right? Or something like that, that can also be the filter to everyone, or all those calls and websites that want to trick us.*

In this strategy, participants added an extra layer that enables protection in multiple ways by filtering, blocking, or blinding what participants hear, see, and receive — from content or people. There is something between the manipulation and the user, but how this layer works varies per machine. While some block only specific content that can be personalised, others block everything, sometimes extending the filter outside the platform. Machine “Online Selector” filters content directly from the brain.

(P12) *The brain would activate when we receive a frequency or data we do not want. The brain activates and, at the same time, activates the ears. We carry something that select frequencies. Then, in front of an unknown number the brain activates and closes this [headset] so we cannot hear anything.*

The extent to which this extra layer filter reflects some ambiguity in the sense of agency participants strive for. On the one hand, participants internalise control by choosing and personalising who and what can reach them; they want to be *free to choose*. “Magic Wall” or “Parabolic Antenna” illustrate these strategies that try to gain control. Using “Parabolic Antenna,” P3 would only grant access to communications with a specific code; everything else is filtered:

(P3) *To make sure that people I want to hear is not filtered] we would have a code so it can pass... which is something that I am already doing. I am not doing this [with a magic machine], but for some people, [...] I tell them to call through another means. I mean, I am doing exactly this because I am very overwhelmed.*

On the other hand, paradoxically, they externalise control by outsourcing the filtering to external actors, because of the claimed need to be *free from* external intrusion. Participants highlighted all the filtering features of the “El Cortijo” machine; it would passively work without any input from the user.

(P6) *Of course, with something next to the router or where the signals are coming, and how it's a magic machine, and it's very smart, and it has artificial intelligence and knows, it's capable of deciding what web pages and calls should filter, so they do not reach me.*

There is an implicit idea of trust in this idea of externalising control. As we saw earlier, our older adults participants feel part of the group of “those who do not know.” Considering themselves as outsiders, they mentioned the need for others to “program” their machines and filtering options. P18 explained that the machine might be more prepared than him because he lacked information.

The ambivalence around the notion of control was exemplified when P12 and P15 commented on the similarities between their machines. Discussing the cons of their machine, P12 explained that she would barely use the machine she made because she does not want to leave something external deciding what she consumes.

(P15) I prefer to know [when the filter is activated] ... because otherwise, if I have a permanent filter, well... I don't know if I am in reality, I don't know... Something different would be that I would protect myself at some moments, but not all the time.

(Facilitator) When would you not do it?

(P15) Mostly, all the time, I would not do it.

(Facilitator) Wouldn't you protect yourself, you mean?

(P15) I mean, I would, when I am going to see something, to explore, but not with other stuff. Only in those moments or unknown calls or... I don't know.

For several machines externalising control, features were nevertheless implemented for the participants to keep an *awareness* of the attempts of manipulation: the feathers on “El Cortijo” would move to subtly notify its users that the filter is in use. This strategy thus elicits a nuanced and ambivalent view on the *need to control*.

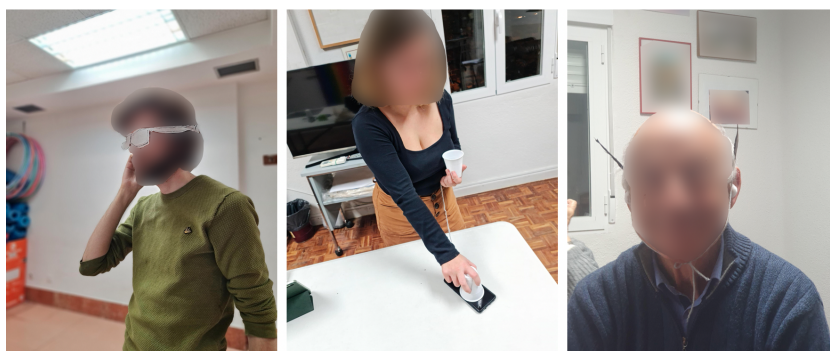


Figure 8.5: From left to right. “The Blind,” “Magic Between Two,” and “Online Selector.” The first two are facilitators emulating the use of the machines, as participants showed during the discussion.

I want to replace the technology

In this strategy, participants expressed a wish to change the technology itself. Participants did not consider the scenario of manipulation, the system, and the manipulator. They created an alternative and simple solution to achieve their goal without external interference. “Magic between two” is a simple children’s string telephone connecting two cups with a string. When participants presented the machine, they justified themselves as if their machine was not good enough but explained the rationale of going back to basic technologies that are less prone to manipulation. Their machine ultimately evolved during the participants’ discourse to add a

filter feature. Similarly, P12, P8, and P31 conceptualise their machine as a headset, and P22 and P33 as a pair of glasses (see Figure 8.5). This strategy relates to *the need not to adapt to technology*, as part of participant's "*right to sanctuary*."

8.5 Discussion

In this section, we discuss the insights gained into older adults' experiences and needs when they face online manipulation. These are essential for the understanding of manipulative designs effects on specific populations like older adults and potential design interventions. The rather abstract imaginary of manipulation that older adults presented during the sessions has an impact on their needs and how these are manifested. It is also likely to limit their actions and sense of agency towards manipulation.

8.5.1 Online Manipulation as The Blob: Awareness of Manipulative Designs in Older Adults

Our participants' conceptions around manipulation tend to differ from previous studies on manipulative designs and felt manipulation in which participants could discern consequences, impact, and make judgments over some manipulative strategies when exposed to them. In an online experiment, Bongard-Blanchy et al. [45] presented nine web user interfaces to 406 participants and asked them to identify potential manipulative designs. Findings show that older participants (despite being sampled from a pool of crowdworkers) identified on average a lower number of manipulative elements than younger participants. Prompted by open-ended questions about manipulation, Gray, Chen et al. [163]'s participants (N=169) could relate manipulation to specific products and designs. More importantly, they distinguished between entities, platforms, and even designers as separate entities in the infrastructure of online manipulation. Our participants' discourse reflected a view of manipulation as an abstract entity. It also reflected associated feelings of powerlessness often expressed during the sessions and the underlying ambivalent need for control. They want to be in control, but are also more likely to outsource decisions, trusting others "who know more about technology."

The metaphors have been useful in some domains to illustrate the mental models of participants or their imaginaries [8] — e.g., "black-box systems" in the field of usable security. While not perfect, inspired by Pitkin's metaphor [362], we relate our participants' experiences of manipulation and corresponding strategies to protect themselves to the metaphor of "The Blob," a sci-fi monster who attacks people unexpectedly in their daily lives. The blob is fluid and has no clear boundaries, it is unclear what motivates it nor how to resist it. These specific characteristics of the blob can support our reflections. Disentangling the Blob is not an easy task, nor one that should fall on the shoulders of users.

Participants shared common feelings of powerlessness and belonging to a group that *needs to catch up*. From previous studies on technology use among older adults [75], we relate these feelings to lower levels of self-efficacy to protect oneself from online manipulation.

Van Deursen and Helsper [449] linked individuals' education level to certain usages of technology. In addition, they found that older females reported less confidence in using digital services. From a phenomenological perspective, Barros Pena et al. [24] explained how older adults lose sense of competence. The effects of individual and situated factors on technological use reported in previous studies align with the experience of online manipulation in our study, involving a majority of lesser-educated women.

The reflection of Zhao et al. [477] about the Continuity Theory applied to older adults echoes our observations. As the authors explain, with age, adults strive to maintain their routines with the knowledge and skills they have. This is important for the ecosystems of online manipulation, which rapidly evolve and, as shown in our sessions, can create a feeling of powerlessness. Drawing from social gerontology perspectives [29], we hypothesise that the case of online manipulation exacerbates this feeling of not belonging, and ultimately the distinction between aging and age, making older adults feel more aged than they are not because of this need for "continuity." Thus, continuity would explain the manifestation of the right of sanctuary and the use of strategies by our participants. Users reclaiming the right to sanctuary, control, and knowledge, illustrate the harmful effects of online manipulation, sometimes leading to exclusion. Yet, they have the right to avoid an extra burden in the digital space. In the next section, we will discuss why this is an important consideration for interventions to fight manipulative design on the user's side.

8.5.2 The Right to Sanctuary as a Response to Manipulative Design: Challenges for Design Interventions

Participants claim the right to sanctuary as not being disturbed and reached, but they want to do it on their terms by controlling who does it or how. These observations resonate with theories of privacy as a dialectical process of boundaries in which users decide what they want to disclose and to whom [221]. This dialectical process can be applied to the experience of manipulation: users decide "what can reach them and what cannot." Gray et al. [169] hinted at this idea of control in their first conceptualisation of manipulative design, but, to the best of our knowledge, this link had not yet been supported by empirical research. Although control has been contested in theories of privacy [329], we discuss a nuanced understanding of how older adults understand online manipulation from their lived experiences. Thus, from the user's viewpoint, manipulative design crosses legitimate users' boundaries and breaks the dialectical process between what the users want and allow from the interface. How older adults can gain control over their interaction is a critical point to consider to designing countermeasures to manipulative design, especially when this population likely shows an ambivalent feeling of agency towards online manipulation. This means that studies investigating design for agency [281] become especially relevant to the design of countermeasures to manipulative designs for vulnerable populations.

In response to powerlessness or low self-efficacy, older adults seemed to rely on external parties, outsourcing control to a third party (human or technological), which they trust

more than themselves for performing online. The strategies and underlying needs we report resonate with prior studies on older adults. In our study and the one of Zhao et al. [477], simplicity is an underlying value of older adults. Moreover, older adults' need for "technology proactivity" echoes our dichotomy of internal vs. external control. In the context of online manipulation, that "proactivity" might challenge users' agency. "Proactivity" and "agency loss" can be seen as two sides of the same coin and should be carefully assessed when understanding older adults' needs.

These reflections have implications for the design of interventions to resist manipulation. Prior studies on manipulative designs on general populations hint at the idea that increasing risk awareness through clues on the interface is effective in counteracting manipulation [38]. Similarly, Bongard-Blanchy et al. [45], Moser et al. [316], and Lukoff et al. [281] discuss the idea of friction as a countermeasure, which was deemed effective in the controlled study of Zac et al. [472]. However, despite older adults' abstract conceptualisation of online manipulation, designing interventions that raise awareness by adding an extra burden on users goes against the right to sanctuary. To think about interventions against manipulative design that work for older adults, we must consider the effect perceived risks have on older adults' technological disengagement [75].

If older adults tend to disengage from technology when they perceive risks, how might we design interventions to counteract manipulative designs without scaring them away from technology? This posits a new challenge in designing countermeasures for manipulative designs for older adults, away from a one-size-fits-all approach for all users. Some lessons learned can be taken from the usable privacy and security domain, in which similar challenges are addressed [117].

8.5.3 What Can We Learn about Vulnerability towards Manipulative Designs?

We started this study with the aim of disentangling what vulnerability towards manipulative design means by understanding how some specific groups are considered vulnerable when they experience manipulative designs. The perceptions of older adults, and influencing factors that underlie their experiences, allow discussing factors that contribute to vulnerability towards manipulative designs, beyond the population of "older adults." We aim to open discussions about what vulnerability towards online manipulation means from an empirical perspective. This will not only be useful for designing interventions, but also for defining and implementing policy interventions that address vulnerabilities online.

Hence, the abstract conceptualisation of online manipulation, which we represented by the metaphor of the blob, might be fueled by other human factors unrelated to aging. Education [45, 280], digital literacy [196, 449, 443], familiarity with certain technologies, or social support [75], all can potentially lead to similar imaginaries and consequences, in line with DiPaola and Calo [113] theorisation of socio-digital online vulnerability. Intersectionality — as a framework in which several axes of oppression intersect [98] —, exacerbates potential harms

and likely sets the tone of some of the insights gained during our workshops. The result is that users with seemingly very different layers of vulnerability may share experiences of vulnerability when interacting with technology.

We therefore identify the following design challenges for the HCI and design community. Even though we started with the category of older adults, we envision these challenges can solve problems of vulnerable users that identify with the experiences of our sample group, regardless of whether they are older adults or not.

- How do we design countermeasures to manipulative designs for users with an ambivalent feeling of agency towards technology?
- How can we design countermeasures to manipulative designs that respect users' right to sanctuary (the right not having to engage, not having to know) without taking away their opportunity to engage with and learn about technology when they want to?
- How to include countermeasures to manipulative designs that raise awareness without triggering technological disengagement in users with low self-efficacy?

While interdisciplinarity is needed to tackle the huge challenges related to online manipulation, HCI as a discipline has a huge role to play in supporting individuals, especially those prone to vulnerability, in resisting manipulation. Platform affordances and design elements are crucial for users to understand, interact, and protect themselves against manipulative designs. This study opens new perspectives and calls for listening to the specific needs of all kinds of populations in situations of vulnerability in the domain of online manipulation to design appropriate interventions.

8.5.4 Methodological Insights and Limitations

This study has limitations that we have considered and tried to mitigate. As a perk of the method, we asked participants to build something to protect themselves. Hence, we refrained from making claims about the machine per se as an agent of protection and carefully assessed if this bias could be reflected in the interpretation of the results. When we refer to the strategies imagined by participants, we focus on the constructed rationale behind the machine's existence, as an embodiment of participants' needs and insights into their ecologies. In some cases, participants' reasoning suggests an overlap between usability-related aspects and the lived experience of manipulation, which we interpret as part of the ecologies of this population, and a starting point to understand what vulnerability towards manipulation means in this context. On another note, we used a perceived self-reported scale to gather the socioeconomic status of participants. This might tend an overestimation since the point of reference of this sample could be lower than the reality.

One could argue that prompting participants by briefly mentioning examples of manipulative situations may impact their narratives. Building on the literature on manipulative designs, which acknowledges the problem of recognising online manipulation, we decided to take a

harm-based approach. We, therefore, prompted situations of harm to the participants — financial loss, privacy loss, or autonomy loss [181] — so they could relate more with their experiences of manipulation as a familiarisation step before building the machine. While we took this decision with an awareness of potential trade-offs implied, participants did not choose scenarios coming from our prompt but drew inspiration from their own experiences of manipulation.

We carefully pondered the decision to conduct the workshops in pairs. Andersen's [10] method aims to be "performative," in which the participant journeys from their internal reflections to the performative task of improvising their machine. With duos, we risked hindering this internal monologue. We believe that these reflections were not hindered, but rather redirected through another person, as a mirror of the experience. While the intimacy between a participant and their machine fades, this conscious trade-off seemed suited to the targeted population, for which we wished to design a comfortable and safe workshop experience. As explained in section 8.3.4, our sample was presumed to be at an intersection of vulnerabilities that could lead to some participants being insecure about making alone — e.g. participants might not be comfortable using their hands for cutting, or show insecurities about 'how little they know' and feel intimidated by the craft material.

We observed participants felt the magic machine method as a fun but activist experience. While the general tone was relaxed and fun, participants took the opportunity in a very serious way to actively protest about how they felt in relation to technology and how little attention there is on older adults. This was visible in some of the results and the conversations during the activity. Some participants felt the creation of magic machines was childish, and joked about it. Although that is a legitimate way of appropriating the activity, some took time to realise we were genuinely interested in the machines. Two participants made a quick machine because they were expecting another activity and two did not make a machine because they felt uncomfortable with the activity. One participant did not manage to embrace the "magic" of the method, and got stuck because she felt "not good with technology." Another participant, uncomfortable creating a machine because of tremor, struggled to express what she would design. Despite this, these participants actively contributed during conversations, and we were subsequently able to also incorporate their views into the results.

We believe the activity was successful and we managed to convey our mission — including older adults in the discourse of online manipulation — to our participants through the different measures of care we took. Still, as some things worked out different than expected, we at times wondered if we did not "care" in the same terms as our participants. We therefore encourage future researchers to make informed methodological choices when applying this method to an abstract topic, and reflect on care by reading the work of Krüger et al. [249, 248].

8.5.5 Future Work

With this study, we open new perspectives for a wide range of future work in the space of manipulative design and online manipulation, as well as empirical approaches of vulnerabil-

ity towards online manipulation. While the present work is a starting point for understanding the experience of manipulation in older adults, we call for a more nuanced understanding in line with previous studies that call for understanding older adults in their real interactions and contexts as "situated communities" [372] and not as an age group. Hence, the interaction with manipulative designs is context-dependent — different contexts have different trade-offs —, we therefore need more phenomenological approaches that understand the lived experiences of older adults with manipulation in specific contexts. We invite scholars on manipulative design to dig into the effects of social ecologies on users in their relationship with online manipulation.

The selection of the magic machine method in our study was motivated, among others, by the need to understand the experience of a phenomenon that works subtly and users might not be aware of: online manipulation. We see opportunities for expanding and rethinking the toolbox of methods that scholars researching manipulative design use to investigate this highly relevant societal topic. Finally, we encourage the community of researchers on manipulative design to work with a broader range of collectives to enhance our understanding of the experiences of online manipulation beyond convenience sampling, ensuring the inclusion of hard-to-reach populations.

8.6 Conclusion

We have conducted three "magic machines" workshops with older adults (N=31) in order to understand their experiences with online manipulation. Through their felt manipulation we have elicited their needs and position towards manipulative designs. We discussed the implications of their needs for the field of manipulative designs, by offering design challenges to consider in the design of countermeasures for older adults. We also reflect on the factors that might contribute to vulnerability towards online manipulation. We offer new spaces for future research on the empirical side of vulnerable populations and manipulative designs.

8.7 Chapter Takeaways

By looking at the experiences of older adults with manipulation in the online context, this chapter provides a more nuanced understanding of what the problem of agency and self-efficacy against these designs means for driving vulnerability, especially when it comes to designing countermeasures. It also reinforces the idea of how social ecology gives rise to drivers of vulnerability.

Perceptions of Manipulation and Resistance Among Low Digitally Skilled Users

This chapter is based on the following publication:

Lorena Sánchez Chamorro. Perceptions of Manipulation and Resistance Among Low Digitally Skilled Users. 2024. Submitted for journal publication.

Abstract. Manipulative designs—so-called “dark patterns”—as ubiquitous design practices that may cause harm online are raising awareness among scholars and policymakers. While academia shows an increased interest in designing counter-interventions, the impact of these designs on users with low levels of digital skills remains understudied. I use an interview study to understand how users with low levels of digital skills perceive and resist manipulative designs (N=19). The findings explain how participants normalise manipulative designs, as well as how socioeconomic conditions mediate the harms coming from these designs. Lastly, I discuss the relationship between digital skills and vulnerability towards manipulative designs and provide directions for future work in the realm of manipulative designs.

9.1 Introduction

How many times a day do you have to reject a cookie consent banner? And how many times do you have to do it on the same webpage? How often have you added one more item to get free shipping, although you knew you did not need it? The use of manipulative designs that try to manipulate, deceive or coerce users into decisions they would not make if fully informed — the so-called dark patterns¹ — is omnipresent in digital services [45, 308], mobile applications [180, 111, 341], videogames [473] and cookie consent banners [38, 183]. This is problematic when we consider its association with different harms to users [341, 181, 382]. The subtle way these types of influence work, like “manipulation” that tries to subvert users’ vulnerabilities in a hidden way, can make it hard to identify for users [45, 287]. As a consequence, the potential harms manipulative designs cause have raised the attention of policymakers [341, 41], and scholars of many disciplines, aiming to better understand to understand the problem and develop counter-interventions [38, 388, 183].

While academic research on manipulative designs has increased in recent years [171, 76], only a marginal body of literature has dug into how personal or situational conditions may render users more prone to be affected by manipulative designs. Some studies explored the role of age in broad terms [45, 15, 382, 383], and some offered preliminary findings regarding socio-economic variables [280, 45]. Hence, the works of Bongard-Blanchy et al. [45] and Luguri and Strahilevitz [280] provided the first insights into how levels of education matter when it comes to interacting with dark patterns: people with lower levels of education are less aware [45], and more prone to engaged with these designs unwillingly [280]. Users’ digital skills level can determine how users interact with manipulative designs and, as a consequence, influence their potential vulnerability [341, 340, 375]. Beyond these initial studies, there is a strong need to understand the extent to which vulnerable groups might interact differently with these designs and how.

Vulnerability is a contested concept, discussed amongst others in legal theories [288, 283, 136, 194]. In this paper, we understand users’ online vulnerability as a situation of power imbalance, in which users are more susceptible to be impacted in the online domain and less likely to recover from harm [375]. Online vulnerability is relational, situated and layered [283, 288]: while everyone can be vulnerable, situated conditions make some users more vulnerable than others [375]. Thus, the extent to which some users are more vulnerable to manipulative designs is not only part of a debate to improve policymaking [41, 340] but also an important question for the HCI and CSCW community, for instance, in the design of counter-interventions [375]. To date, a major need around to question of online vulnerability is to study a diversity of populations that have not been included yet in the conversation around manipulative designs.

This paper aims to document the experiences of populations that are traditionally considered vulnerable to manipulative designs — e.g., users with lower socio-economical status

¹The research community is studying this phenomenon using a variety of labels, including deceptive design, nudges, anti-patterns, and most dominantly, “dark patterns.” Following the ACM recommendations on diversity and inclusion [143] I hereby use the term “manipulative designs” to describe this phenomenon.

(SES), or minorities [136, 283] —, and I operationalise this idea of vulnerability through the construct of digital skills. As socio-economic factors are a known predictor of digital skills [195, 443], understanding how users with lower levels of digital skills interact and resist manipulative designs can provide insights into how socio-economical factors play a role within these design practices. While embracing the critics against labelling specific populations as vulnerable, which can be stigmatising [136, 375], this study aims to understand how manipulative designs operate on specific populations whose experiences and ecologies can be transferred to other populations, and ultimately to detach vulnerability from the specific labels.

To understand how vulnerability operates in the context of manipulative designs, I have conducted an interview study with users in situations of vulnerability (n=19). Participants were recruited in collaboration with an NGO (Non-Governmental Organisation) that works with populations at risk of social exclusion. The study consisted of three parts. First, participants were asked to perform simple tasks on two websites including manipulative designs and asked questions about these. Second, building on the critical incident technique, I conducted a semi-structured interview about their experience of online manipulation and mechanisms of resistance to it. Third, the interview was finalised by showing manipulative designs to the participants and gathering their experiences when they encountered such strategies.

The contributions of this paper are manifold. First, by explaining the experiences with manipulative designs among users with lower digital skills, this paper offers a more nuanced understanding of the effects of manipulative designs on populations in situations of vulnerability. Second, this study discusses the results in light of existing design knowledge, articulating challenges for the HCI and CSCW community to design counter-interventions. By reflecting on these experiences, this paper deepens into the idea of agency, resistance and self-efficacy among low digital skill users and contributes to unveiling what resistance means for users in situations of vulnerability. I, hence, invite scholars and practitioners to rethink and design countermeasures to manipulative designs that account for such vulnerability.

9.2 Related Work

9.2.1 The Experience with Manipulative Designs and Online Vulnerability

Vulnerability is a multifaceted concept that has been primarily discussed in legal theory [283, 288, 136], and that refers to a position of higher susceptibility in which the user is less likely to recover [340, 288]. While the research at the intersection of manipulative designs and vulnerability is limited [375, 340, 341], the extent to which users are vulnerable to manipulative designs is a rising concern among policymakers [341, 340, 41] given the associated harms. Research has indeed demonstrated a variety of harms caused by manipulative designs, for instance, cognitive, financial, privacy-related, or linked to identity and misinformation [181, 382]. However, as shown by Bongard-Blanchy et al. [45], even when users can recognise some manipulative designs when exposed to them, they often fail to link the designs to underlying

harms. Sánchez Chamorro et al. [382], for instance, showed the role played by peers and family members for teenagers to see the harm caused to them through their interaction with manipulative designs.

While scholars and policymakers have sometimes labelled specific categories of users as "vulnerable" — children, older adults, minorities — [41, 340], we here follow the idea of layered vulnerability: all users are vulnerable, but some are more vulnerable than others. Luna describes layers as different users' conditions — e.g. personal, contextual, physical — that accumulate and overlap, providing a continuum of vulnerability, beyond a user category. Malgieri [288] introduces the notion of drivers of vulnerability as triggers to these layers in order to explain what makes users vulnerable in the digital realm. Investigating a variety of users' experiences with manipulative designs will thus help the research community to understand the contexts and ecologies that drive vulnerability to manipulative designs [382, 383]. In other words, what makes them vulnerable to manipulative designs. While embracing the critiques towards using "vulnerable" as a user category, this study focuses on users with lesser digital skills as part of risk at social exclusion collectives, aiming to be a starting point to disentangle factors of that vulnerability. By understanding the context of this population we can elicit challenges to make countermeasures for all populations who share similar needs and ecologies and experiences.

Manipulative designs are relational: users only perceive them when they receive an impact [382, 383, 163]. Thus, Gray et al. [163] reflected on the idea of "temporality" of manipulative designs. To understand users' relationship with manipulative designs, I look at the experience of manipulation as a proxy, inspired by Gray et al. [163] since the idea of felt manipulation and deception as a proxy is tied to the relational aspect of manipulative designs.

9.2.2 The Experience of Manipulative Design Practices

The literature on manipulative designs has expanded in recent years [171], but is still limited in the understanding of how users interact, cope with manipulative designs, and in which contexts these reactions happen [171, 381].

A limited set of literature has explored how different conditions affect the relationship between users and manipulative designs [381, 171], like age [382, 45, 383, 15], education [45, 280, 472], socio-economic status [472], or used device [447]. The works of Bongard-Blanchy et al. [45] and Luguri and Strahilevitz [280] give the first intuition of how the education levels and socio-economic status matter when it comes to interacting with manipulative designs. Bongard-Blanchy et al. [45], via an online survey study, presented how users with lower levels of education showed less awareness of manipulative designs when exposed to them. Luguri and Strahilevitz [280], with an experimental setting in which users were asked to conduct tasks on websites with manipulative designs, showed how users with lower educational levels were more prone to unwillingly be engaged in these designs [280]. These studies resonate with the publications from the OECD [341], aligned with the idea of socio-digital vulnerability [113]. It is important to note that Zac et al. [472], via an experimental study, do not reflect the same trend of results. They do not find a relationship between education and socioeconomic status with

susceptibility to manipulative designs. Notwithstanding, in line with Dechant et al. [106], who calls for debunking vulnerability in videogames by looking at the harms that come from their design, we believe more research that focuses on the experiences with manipulative designs to get a more nuanced understanding of the contexts in which this happens.

From an experiential point of view, Bongard-Blanchy et al. [45] observed a relationship between users's perceived self-efficacy in resisting manipulative designs and their capacity to recognise them. Similarly, when Maier and Harr [287] showed manipulative designs to undergraduate students, they observed participants assessed these designs in relation to their potential harm, while justifying their existence in some cases. Some reasons that Maier and Harr's participants gave to accept nagging techniques — i.e. design patterns that insist on asking the user about the same issue [174] — or forced action — i.e. forcing the user into something — are their visibility and the possibility of choosing to get out of them by “closing the page”. Building on this kind of user' assessment, Gray et al. [163] observed the temporality of the experience of manipulation in the presence of these designs. From an initial judgment to negative results from the interaction, there are different stages in which users perceive they are being manipulated through the interface that carries negative experiences and emotions — distress, upset, guilt, fear, hostility, irritability or shame, among others — in line the results of Avolicino et al. [15]. Indeed, M. Bhoot et al. [286] found, in a survey-design study that showed manipulative designs, that frustration correlated with the identification of manipulative designs.

Some studies on manipulative designs have highlighted the role of the perceived agency of users [281, 285, 366, 78]. Schaffner et al. [390], when investigating possible countermeasures to manipulative designs, discuss the ambivalent phenomenon of users wanting to remove information that might confuse them, while simultaneously being made aware of what is removed: they demanded more agency of the tested countermeasure and resonating with the results of Owens et al. [347].

9.2.3 The Third Digital Divide and Resisting Manipulative Designs

The relationship between digital skills and online manipulation relates to the conception of socio-digital inequalities. Socio-digital inequalities are the differences in the acquisition, use and outcomes that users can get out of technology that is determined by socio-economic aspects like education or socio-economic status [195, 196]. They are interrelated, thus, while access to technology is important to make use of it, specific digital skills — e.g. use of social media, use of services, or coding — would contribute to different outcomes — e.g., from buying an online ticket for the cinema to obtaining specific jobs [446]. Therefore, the level of skills might be a determinant of protecting oneself from manipulative designs as an outcome.

Considering digital skills becomes especially relevant to understanding resistance to manipulative designs. While “resisting” is a term that is used in the manipulative designs research [45], there is no actual definition, and it is commonly understood as users not falling into the manipulative designs. In this paper, I understand resistance as the different strategies that users enact to protect themselves from manipulative designs or their potential effects. I,

therefore, get inspiration from protection theories in psychology, which establish that to protect themselves, users need to perceive an appraisal — Protection Motivation Theory [404] —, know about the intention of persuasion — Knowledge Persuasion Model [130] — or, be aware of the risk to develop a coping strategy [344]. With this rationale, several studies on manipulative designs argue about educating users about manipulative strategies [45], or develop interventions to increase knowledge about the strategy when the interaction is taking place [80].

While manipulation resistance is still understudied, prior studies provide insights into the usefulness of focusing on resistance strategies. Through their online survey study, Bongard-Blanchy et al. [45] explained how people who recognised more manipulative designs reported higher self-efficacy to be manipulated. On the contrary, Strycharz et al. [427] demonstrated in a controlled study that increasing the knowledge of participants fell short in helping them to use opt-out options for personalisation on online services. Similarly, Klütsch et al. [237] found that providing knowledge about the manipulative design during the interaction with cookie consent banners did not relate to less consent.

Knowing the experiences in which users with lower levels of digital skills feel manipulated and resist that manipulation can support the development of interventions that foster those strategies. If theories explain that users protect themselves when they are aware, but manipulation works subtly, we need to see when and how they are aware, to design interventions to trigger and incentivise the coping strategies that best work for them. In light of the aforementioned work, this paper addresses the following research questions:

- (RQ1) How do young adults with lesser digital skills experience manipulation?
- (RQ2) How do young adults with lesser digital skills experience harm coming from manipulative design strategies?
- (RQ3) What are the mechanisms that low digitally skilled users use to resist manipulation?

9.3 Methodology

To address the research questions, we conducted an interview study composed of three parts: interactive tasks, semi-structured interviews based on the critical incident technique, and the use of probes including manipulative designs. Before conducting the study, ethical approval from the University of [anon] was granted.

9.3.1 Participants and Recruitment

Participants were recruited in collaboration with an NGO conducting socio-cultural interventions with populations at risk of social exclusion in [anon]. After contacting the NGO and clarifying the study's implications, the organisation members promoted the study among two par-

participant groups. I make the distinction in Table 9.1 for the sake of contextualising participants. This table can be found in the Appendix.

1. The first group (here labelled as “A”) was composed of 11 participants who belonged to a program in which participants were trained to become administrative assistants.
2. The second group (here labelled as “B”) consisted of 8 participants who belonged to a training program in gardening skills accompanied by upskilling activities around job-seeking and daily life (e.g., managing their relationships with public administration or using Microsoft Office).

Overall, both groups were diverse. Participants’ ages ranged between 18 and 28, and they came from a diversity of countries. While they differed, most of the participants were foreigners who joined these programs in pursuit of upskilling and helping them find their way. The psychologists who work on the NGO premises helped me contextualise the participants’ situation: participants often lacked a support network, which made them more distrustful overall.

Flyers to promote the study were placed at the venue, and I gave an in-person introductory talk to present the study before the sessions started. The NGO provided a room to conduct the interviews where study participants could come and go at their convenience. Once participants entered the room, they were reminded about their rights and gave informed consent.

9.3.2 Protocol Overview

This study was conducted between the 21st of March and the 18th of April of 2023, at the NGO premises. I conducted a 3-steps qualitative study, triangulating different data types. Using websites with manipulative designs, I invited participants to conduct two tasks. This was followed by a semi-structured interview using the critical incidents technique. Last, by showing examples of manipulative designs via static images of platforms, I qualitatively explored participants’ insights and experiences about these examples.

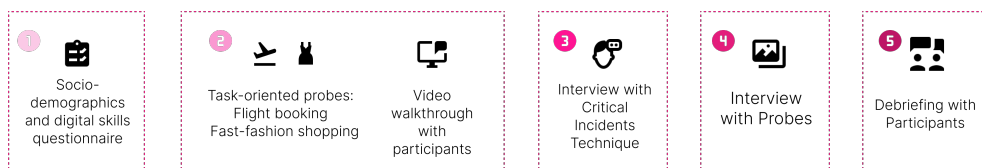


Figure 9.1: Main steps of the study protocol

1. *Socio-demographics and Internet skills questionnaire* To check participants’ digital skills level, I used the Internet Skills Scale [444]. I used a translation into Spanish provided by the Disto project [anon] [120]. I adapted the questionnaire to the norm and usage of Spanish in [anon], which uses different terms than other Spanish-speaking countries

(e.g. “móvil” instead of “celular”). Participants filled out the questionnaire in presence of the researcher so they could ask questions if needed. This questionnaire was meant to have an overview of how participants found themselves according to different domains of digital skills: operational (i.e. “technical competences required to command a computer” [446] (p.6), informational (i.e. skills related to information search), social (i.e. competences related to social interactions and disclosure online), creative (i.e. skills related to the creation of content), and mobile skills (i.e. abilities with mobile phones and their platforms). An overview of their skills can be found in the Appendix.

2. *Interaction Tasks*. Participants were asked to conduct two tasks. First, they were asked to reserve a flight on the airline company website “Ryanair.com”. Second, they were asked to go through the purchase process on the e-commerce website “Sheln.com”.

While participants conducted the tasks, the screen was recorded. Some participants thought out loud — although it was not required —, which was used as data. After finishing the tasks, participants were asked about their decisions during the process and their encounters with manipulative designs.

3. *Critical Incidents Semi-structured Interview*. In this part of the interview, participants were asked to remember negative experiences in their Internet usage, progressively laddering into experiences of deception and experiences of online manipulation. The topics covered in the interview were: internet usage, moments in which participants had bad experiences online, perceptions of felt deception, perceptions of felt manipulation, and responses to those incidents. In some cases, if participants did not mention any harm, two questions were asked about excessive expenses and time investment online. This technique was meant to ask for experiences of manipulation in real contexts to unveil how the interaction occurs.
4. *Manipulative Design Patterns as Probes*. In the last part, participants were shown different manipulative designs that were meant to act as probes [44]. Participants were asked questions about them in relation to their encounters, their perceptions of potential harm to users, and their reactions towards them. They were shown as patterns, through screenshots, simultaneously on a digital board. As participants reacted to the probes, I followed up with multiple “why” questions to elicit participants’ awareness, reactions and general interaction with these designs in the users’ daily routines. I then asked “What do you do when you encounter such an interface?” to understand the behavior and potential vulnerability to the manipulative element and “What happens after?” to understand the potential harms and coping strategies.
5. *Debriefing*. Participants were explained about the manipulative designs they had seen in the interview and given an informative brochure about online safety.

ID	Group	Age	Gender	Country of origin	Education	SES	Device
1	A	26	Female	Colombia	Secondary education (A-levels)	3	Laptop
2	A	19	Female	Polonia	Secondary Education	7	Mobile phone
3	B	19	Female	Spain	Primary education	6	Mobile phone
4	B	21	Male	Mexico	Secondary Education (A-levels)	8	Mobile phone
5	A	24	Male	-	Secondary education (A-levels)	4	Mobile phone
6	B	19	Female	Spain	Primary education	3	Mobile phone
7	B	21	Male	Morocco	Primary education	5	Mobile phone
8	B	27	Female	Bangladesh	Secondary education (A-levels)	5	Mobile phone
9	B	23	Female	Romania	Primary education	4	Mobile phone
10	A	17	Female	Spain	Secondary education	6	Mobile phone
11	B	24	Male	Venezuela	Secondary education	4	Mobile phone
12	B	23	Female	Equatorial Guinea	Secondary education	4	Mobile phone
13	A	23	Male	Bolivia	Secondary education	5	Mobile phone
14	A	17	Male	Spain	Secondary education	6	Mobile phone
15	A	24	Female	Venezuela	Secondary education (A-levels)	7	Mobile phone
16	A	21	Female	Venezuela	Applied Bachelors	7	Mobile phone
17	A	28	Female	Dominican Republic	Primary education	3	Mobile phone
18	A	18	Female	-	Secondary education	3	Mobile phone
19	A	21	Female	Honduras	Secondary education (A-levels)	5	Mobile phone

Table 9.1: Description of self-reported demographic data. More information can be found in Appendix. For self-reported education, participants were asked to mention the highest level of education achieved. Participants rated their self-reported socio-economic status (SES) on a socio-economic scale, from 1 (lowest) to 10 (highest).



Figure 9.2: Sketch of the e-commerce website with choice architecture manipulation shown to participants.

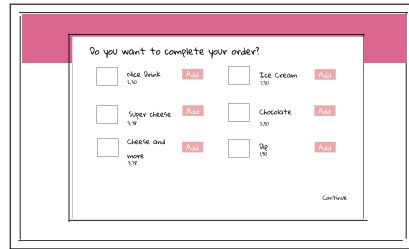


Figure 9.3: Sketch of the website to buy pizza with nagging and choice architecture manipulation techniques shown to participants.

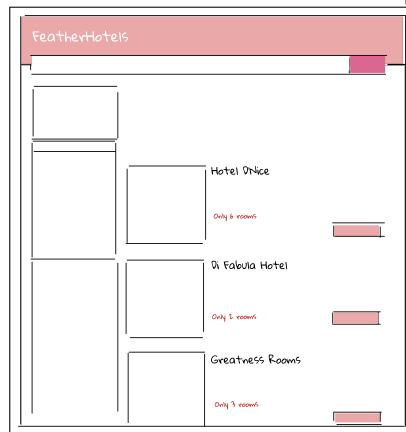


Figure 9.4: Sketch of the website to reserve hotels with scarcity claims shown to participants.

9.3.3 Data Analysis

The interviews were audio-recorded and transcribed. For the coding, I followed an inductive-deductive approach with different coding strategies, as explained in [378]. I primarily used three types of coding. *In vivo* coding captures literal words as reproduced by participants, which is relevant to give voice to participants. The way they express this topic says a lot about their perceptions. For instance, participants referred to themselves as "addicts", and although they did not approach it in a clinical term, their vocabulary encapsulates relevant meaning. *Descriptive* coding summarises a phenomenon with the words of the researcher. Therefore, it can embody different lenses (e.g., a participant gave their personal data; she has been afraid of online services). *Versus* coding summarises a phenomenon through the opposition of two concepts. The researcher understands what participants are describing always with a confrontation (e.g. THE PLATFORMS VS ME, THE INFLUENCERS VS THE INFLUENCED). Given the nature of the topic we discussed — sometimes an abstract entity of which participants are not

expected to be aware or have knowledge —, it proved to be useful to give some nuances to the experiences of manipulation.

To analyse the data, I used an interpretative reflexive thematic analysis process, following Braun and Clark approach, as a way of identifying patterns in meaning within the codes [51, 50]. First, I familiarised myself with the data and coded with the aforementioned strategies. After coding, I created preliminary themes based on the analytical memos taken, in which first reflections about similitudes and differences between participants, as well as connections with theory and literature, were made. Starting with these main themes, I analysed the codes, looked for patterns, and created initial themes and sub-themes. I discussed the data and initial themes with two fellow HCI researchers and, after several iterations, I defined the final themes and sub-themes that addressed the research questions. I supported this analysis with conceptual maps to visualise codes and themes.

Inspired by Sánchez Chamorro et al. [382] analysis, when looking at harms coming from manipulative designs, I used a deductive approach, building on the framework of harms described by Gunawan et al. [181], and complemented it by Sánchez Chamorro et al. [382]. Similarly, I built on the existing ontology of Gray et al. [174] when I looked at manipulative instances, which also include “attention capture deceptive patterns” as one category of meso-level patterns. However, attention capture deceptive patterns combine 11 different design strategies. Thus, in the reporting of these attention-capture deceptive patterns, I use the terms coined by Monge Roffarello et al. [315].

9.3.4 Ethical Implications

Given the specific implications of this study and the situations of the participants, care was a priority when conducting this study, understood as prioritising users’ needs during the study [435]. Specific measures were taken into consideration when recruiting participants, conducting the interviewing process, and debriefing. The recruitment information was provided a month in advance to the NGO so participants were in advance with time enough. Before the interview started, participants were given enough time to read the information, and I additionally explained every point of the consent form.

The interviews were conducted at the NGO premises to facilitate participation. The NGO allocated a room for the interviews, which was continuously available throughout the month, so potential participants could decide when to participate. The purpose of allocating a continuous specific room for the activity was to avoid interrupting their usual activities at the NGO as much as possible. For this reason, I had to accommodate their schedules. For instance, on two occasions, participants had to stop the interview and resume it afterwards. Spending a full month working on the NGO premises allowed me to build rapport with participants.

During the debriefing, I paid special attention to explaining that the images they saw were meant to manipulate and cause harm to users in a stealthy way; therefore, it was not unusual if they did not know them. I explained the mechanisms and prepared a brochure with information about online safety. An informative poster with the debriefing key points and the information for consent withdrawal was provided to the NGO to hang after the interviews.

ID	Operational	Informational	Social	Creative	Mobile
1	6,67	5,20	6,40	5,50	7,00
2	6,67	5,60	7,00	5,00	7,00
3	1,67	4,40	3,00	1,00	4,00
4	5,33	5,60	6,80	6,25	5,67
5	6,00	5,80	6,20	6,25	5,00
6	5,33	6,20	7,00	4,25	5,67
7	5,33	5,40	5,40	5,50	6,67
8	5,33	4,40	4,80	4,00	4,00
9	3,67	2,60	3,80	5,50	6,67
10	5,33	6,00	7,00	3,25	5,33
11	3,67	4,20	4,20	1,75	5,67
12	3,67	4,80	6,60	2,25	6,00
13	5,00	6,80	6,00	5,50	6,67
14	4,67	3,60	4,60	4,00	5,00
15	7,00	5,40	7,00	5,25	6,67
16	6,67	4,80	7,00	5,50	7,00
17	3,67	4,40	4,20	4,50	5,67
18	6,00	3,80	5,00	5,25	7,00
19	7,00	5,20	7,00	4,50	7,00

Table 9.2: Description of self-reported digital skills. The table shows the average for every domain building on a self-reported Likert scale (1-7). More information about the scale can be found in the Appendix.

9.3.5 Positionality Statement

For this study, I embraced my subjectivity as a resource [161, 238], so I find it important to situate myself within the work and reflect on my position. The study of manipulative designs and their relationship with vulnerability is a personal drive that moves me as a researcher. Although I presented myself to the participants as a researcher, I also shared my identity as a neighbour and showed my interest in helping users in their online interactions. Sharing a similar socio-economical, cultural, and geographical background with the study participants, I recognised myself in some of the stories and experiences they shared, and also recognised the reality of the neighbourhood where I grew up. However, as a non-migrant within the country, I also acknowledged that I could not relate to the experiences shared by some of the participants.

Beyond the country of origin, a notable difference between my participants and me is the age, or the stage of life in which we are, which implies differences in the usage of the internet and mobile applications. I am a frequent user of a diversity of platforms that contain manipulative designs — e.g., Twitter, TikTok, YouTube Shorts, e-commerce, and video games —, which were commonly mentioned by participants during the conversations. However, I acknowledge that the way we use it, the contexts, the purpose of use, and, above all, the content consumed might diverge. Being aware of that, I embraced that difference as a resource for my reflective interpretative analysis. By reflecting on those differences, I could articulate key links between manipulative designs and their manipulation experiences.

9.4 Findings

In this section, I report the findings in three main overarching themes that describe important instances of participants' experience of manipulation. First, I explain how the experience of manipulation leads to the normalisation of manipulative practices and the role of trust in that process. Second, I present how contextual and social aspects mediate how participants experience harm. Third, I show resistance mechanisms for manipulative practices that participants use in response to harm.

9.4.1 Imaginaries of Manipulation: the Normalisation of Manipulative Practices

The interviews unveiled a recurring pattern of associating influence with the influencer economy, how platforms can influence people through content, and economic scams online. These experiences have been very useful in understanding the imaginaries of manipulation, harm, and coping mechanisms. In this section, I will focus on insights into the imaginaries of manipulation concerning platform design, which is the main lens of this study.

This theme explains participants' conception of platform design-enabled manipulation [25], which pivots around two ideas: the awareness and the acceptance of manipulation. Par-

participants have notions about how some interfaces try to steer users in different directions, and they experience such intents of manipulation by feeling that online business and platforms “play with your mind” (P10). However, they normalise the phenomenon and accept it as “business as usual”. Thus, participants do not identify as “risky” what for them is normal. This turns into self-blaming: participants put the responsibility on the users for the influence that platforms have on them. Lastly, this theme explains how the role of trust is vital in determining what is “questionable” or not in participants’ online interactions.

Awareness and Acceptability

Participants showed awareness of how the interfaces they see are trying to steer their decisions. From the interactive tasks, where some participants explained they never interact with some elements because it can be “fake” or a “scam”, to the last stage with probes, where participants expressed that manipulation is what you should expect from the internet, this idea of confrontation — “the internet vs. us” — was very present. Expressions such as “they always want to trick you”, “they play with your mind, and you fall (P10)”, “they try to manipulate you with those games because nobody is going to give money for free” (P19), or “it is a stalemate, you click, or you click” were common in the discourse of participants. However, participants tend to perceive these design instances as normal: participants explained how they were used to accepting consent banners or looking for the “X” to close constant pop-ups, for instance. Indeed, it stood out that some participants internalised that interacting with the manipulative design, by accepting some options, was the expected way to continue on the platform.

P9: “[...] There are so many times that, for instance, if you don't accept Google, it does not allow you to get anywhere. And sometimes, I have that mania that as it is popping up, I click 'accept everything' or sometimes it gives you reject option, but I say 'what for?'.... because, no, I don't take those things as something important, so I say 'no, nothing is going to happen,' so I click [...]"

In this quote, P9 explained that he accepts cookie consent banners on Google because otherwise, you cannot access any website, and he does not consider it important because he sees no consequences in accepting them. The absence of apparent consequences seems to create a reinforcement effect in which participants can continue interacting with the manipulative design without consequences. Indeed, P7, who was not fully proficient with the local language, explained he accepted cookies “because you want to continue; otherwise, you are leading to the settings”. It feels he was used to cookie consent banners as a way of guaranteeing that he would remain on the website. As shown in Figure 9.8, this normalisation led P1 to fall for a product subscription offer.

Although participants had experienced the effects of these designs, they did not necessarily recognise them as malicious. Participants normally associated scarcity cues or choice architecture manipulation — changing colours and showing intricate wording — with marketing strategies. Participants even sometimes consider them useful: many participants related some manipulative practices to part of the service — e.g., “They are there to inform you” —, or

to some useful exchange — e.g., *“If I did not have the scarcity cue I would spend more money”*. Although participants recognised the attempts to steer them, they felt rather safe and in control over their interactions in front of these practices. In this way, although the presence of nagging techniques, choice architecture manipulation, and complicated wording that led participants into giving excessive personal data, encountering undesired ads, or buying unexpected products, was common in participants’ experiences, participants generally shared the idea that *“as long as I click X, nothing can affect me”*.

In this regard, participants’ understanding of scarcity cues deserves special attention. For participants, scarcity cues are another way in which companies try to steer users and pressure them, but these designs are not necessarily true for them. Indeed, the general feeling is that scarcity cues are fake: *“as it says there is only one day left, I do not buy it because I feel it might be fake”* (P6). P10 and P12 experienced episodes where they realised that scarcity cues were fake: they would come back to the website and still see the product that was supposed to be scarce.

Resisting Manipulative Design is our Responsibility

There is an entanglement of manipulative practices and interface burdens that participants feel they need to dodge: pop-ups to close, struggles to find the “X”, notifications, small fonts, different colours, hidden sponsors, or texts full of technical language are some of the burdens that participants reported to commonly find online. Although participants have normalised these design features, they feel pushed to dodge these instances because they are omnipresent and seem to take the blame on them if they do not do it. The narrative of not putting enough effort in avoiding these practices was common among participants: *“It is our fault because we are not informed”* (P18), *“people don’t read”* (P6) or *“it is peoples’ fault”* represents the common sense of guilt very well.

The Role of Trust in the Experience of Manipulation

Participants showed awareness of some intents of manipulation or deception online. They declared to be aware that something is deceptive when they see it, but their interactions are built on trust or mistrust towards a system. Trust in services seemed to be the driver in assessing the “benevolence” of a service. There was a general trend to admit accepting cookie banners on Google or Ryanair because they trust these services and believe their data will not be misused. P15 explained accepting cookies *“because the website is trustworthy”, “as it is Google, she would not think much about the cookie acceptance, because Google is safe”*. With a similar rationale, P1 reflects on how cookie consent banners are a constant that people have normalised, and although she — and others — do not know what it is about, she trusts the services, so she normally accepts. Indeed, she even referred to *“my”* Google when talking about the provider.

P1: *Because it says we value your privacy all the time [...] What is more, in everything I get in, and it says ‘we value your privacy’, and you have to accept, because otherwise,*

I'm not interested, I don't know... I find it curious...

Interviewer: Why so?

P1: Because it didn't appear in the past, and now they are everywhere, and it is very normalised, so people accept it and totally accept it. I don't know what I am accepting, but yeah... I didn't even think about 'well, I am going to look at what I am accepting' because of time, and I decided to go directly. I trust my Google. I trust these platforms, therefore I accept because I trust and it is better [do it] automatically.

The extent to which participants trust what they see and the associated service behind it falls under the manipulative interface. But, what is causing mistrust within participants? Within their own experiences of deception and manipulation, and by showing probes, participants expressed different aspects of the platform affordances that triggered mistrust. A (i) *mismatch of expectations* between what they believe to know from the platform and what they perceive, (ii) references from *third parties*, and (iii) *knowing the platform* were some of the triggers of trust or mistrust reported by users.

Participants reported noticing intents of manipulation or deception when they perceived a mismatch of expectations between what they believed to know from the platform and what they saw. P18, for instance, realised her friend was hacked because she realised something was wrong with their style of writing posts and private messages. Similarly, P15 realised through the writing style that some accounts are bots. When seeing the disguised ad in the probes we presented to her, P19 said “*it is just something you do not put in news*”. Thus, although one participant reported that she does not trust sponsored content when she sees it, none of the participants realised the disguised ad was sponsored, and not a news article. Trust in these interfaces is also generated by third parties, such as recommendations from friends and family and referrals. Several participants reported how they would trust a platform if someone recommended it to them. Similarly, it was a strategy to look at the reviews on platforms in order to guarantee the platform was trustworthy.

Being familiar with the platform would also increase participants' trust. Thus, several participants explained how they knew the platform in which the interactivity tasks were performed and, therefore, they would feel comfortable. For instance, P14 explained the first time she went to the platform, she was nervous, but now she was already used to it. Repeated interactions without apparent bad experiences reinforce trust in the platform.

9.4.2 The Socio-technical Architecture of Online Harms

This theme expresses how the experience of harm depends on socio-technical aspects. Technological affordances — notifications, attention deceptive capture patterns, choice architecture manipulation, and algorithmic personalisation — target the social context of participants — socio-economic or personal situation — to reinforce their effects. Context and technological affordances of manipulative designs thus reinforce each other in creating experiences of harm.



Figure 9.5: Sketch representing the Disguised Ad showed to participants

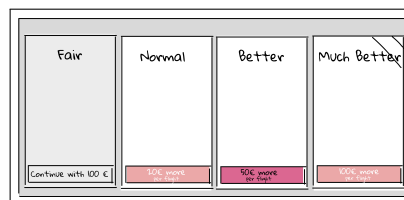


Figure 9.6: Sketch representing the Choice Architecture Manipulation showed to participants

An expression of this phenomenon on the user side is the idea of “tangible harms”: participants identify harms related to their contexts. For instance, financial harms, and attentional or psychological harm (“being addicted to” in their own words). When asked if they feel influenced by platforms, they responded negatively since their idea of influence and manipulation relates to the “influencer economy” or to “financial scams”, and they do not identify as influenced by them. However, when explicitly asked if they had spent more time or money online than expected, they all declared to have done it. There is a tangible aspect tied to financial, attentional, and emotional harms, which are more difficult to find in privacy harms. Table 9.3 summarises the different socio-technical and personal mediators of harms that participants have reported and their link to manipulative designs. This table is not meant to be exhaustive but a supportive visual tool.

Emotional harms

Participants reported experiencing emotional harms derived from people, being intertwined with the attention economy ecosystem. Participants’ discourse shows a general association between the use of mobile phones and emotional harms. Terms like “addiction” and “slaves” were common in the participants’ vocabulary. Cyberbullying, online harassment, and online hate on the one hand, and influencers on the other was reported as a cause of fear from participants, who sometimes see potential physical harm. The influencer economy was reported to be a cause of distress for most participants: seeing some content would trigger inner insecurities and comparison. One participant reported to have experienced eating disorders in the past, thus be very cautious about its relationship with Instagram nowadays. While this is not strictly related to manipulative designs, it is here reported as it also belongs to the participants’ experience of felt manipulation.

Participants reported feeling anxiety, “urge”, and annoyance in the presence of some manipulative patterns like cookie consent banners that nag all the time and they end up accepting (P19, P1, P10). They also felt impulsivity caused by the orchestration of discounts, scarcity cues and the position and size of those (P10) (See Figure 9.7). P6 also explained she felt “addicted” to those platforms and would not stop buying. These harms seem also to be related to financial harms on e-commerce platforms. Participants reported similar anxiety in relation to constant notifications on their phones.

Expressed Harm	Socio-technical and personal mediator	Manipulative design associated
Emotional Harm	Impulsivity Loneliness	Activity messages Infinite scroll* Nagging (M)
Attentional Harms	Loneliness	Activity messages Infinite scroll* Algorithmic personalisation (M)
Financial Harms	Economic situation	Nagging (M) Drip pricing (L) Scarcity cues (M)
Privacy Harms	Perception of consequences	Nagging (M) Choice architecture manipulation (M)

Table 9.3: This table relates the harms reported by the participants to design patterns described in Gray et al. [174] through the socio-technical and personal mediators observed during participants' reported experiences. Designs marked with (M) correspond to meso-level manipulative designs, while those marked with (L) refer to low-level manipulative designs. The designs marked with "*" correspond to attention deceptive design patterns, as gathered in Monge Roffarello et al. [315] This table is not meant to be exhaustive but rather a supporting visual tool to map reported harms with associated manipulative designs

Attentional harms

While most of participants reported not to feel influenced by any platform, they also declared to have spent more than they had planned on social media platforms. While most of them cannot articulate why social media takes hours of their time, others relate it to the content and to some deceptive affordances like notifications, or infinite scroll. P18 explained she thinks *"TikTok is very addictive... although in Instagram I cannot spend a lot of time, it's true I open it a lot to see who talked to me and the notifications, to see if someone has posted something new, the stories, and all of that."*

On the other hand, attentional harms seemed related to participants' social context. Several participants have expressed they spend so much time on their phones because of loneliness, as they use these applications to evade their reality and responsibilities. As explained by the psychologists at the NGO, most of the participants lacked social support at home or were separated from their families. The conversations with the participants made the role of the lack of support in reinforcing harms visible. P9 exemplifies well the use of some platforms as a substitute for social support:

P9: *"I don't know, I am with the phone everyday because I have nobody next to me, for instance, a family member or something so I can stop with the phone. The only thing that gives me life is the mobile phone. You go on TikTok, then on Instagram, then Whatsapp or someone calls you [...] and that's how I spend the day"*

Financial harms

Economic incentives are crucial in how participants understood online manipulation in general, and coming from the designs in particular. Several participants explained how they were deceived online by abusing situations of vulnerability. P7 and P1 fell into giving personal data when they see advertisements about “supposed jobs”. P7 was playing free mobile games when he saw these pop-ups. Similarly, P13 was the victim of an online scam that promised to get into the trials of a football team. P1 receives a lot of content from free courses on Instagram; which leads to her giving a lot of personal data and spamming her mailbox folder. When asked what she thought caught her attention when she subscribed to the course, she said: *“That was for free, that is what caught my attention”*. She also explained how she tried to use the internet to learn as much as possible. The type of content they receive, according to profiling, tries to exploit their economic situation as a factor of vulnerability, which is fundamental to understanding how they experience online harm.

Economic loss was the most perceived harm by participants, who would sometimes associate “influence” with “money”. Thus, in some cases, participants assess whether or not some manipulative designs would affect money loss. For instance, scarcity cues would not affect P12 because she will always value the price most. Similarly, participants explained choice architecture manipulation as problematic because it is inciting to buy. When exposed to the flight purchase task, the participants who wanted to save money were selecting a more expensive option because they did not realise the manipulative design.

The economic situation was also visible within participants' relationship with manipulative designs. When conducting the interactive tasks and showing the associated manipulative probes, participants easily recognised some manipulative designs they usually experience. Frequent users of the e-commerce platforms recognised they were influenced by the discounts that pop up on the website. Beyond such nagging technique, in Figure 9.7, P10 explained the intertwined experience of manipulative designs that use economic incentives on such websites: the combination of economic incentives exploiting the sunk-cost fallacy [296] — as participants are already making an investment, adding something extra is a marginal loss that “compensates” — with the location of such economic incentives at strategic places in the user journey, makes it impossible to resist: *“They normally get you”* (P10).

Privacy harms

The expression of privacy harms also embodies the idea of tangibility. Participants reported experiencing — or not — privacy harms within their use of platforms, directly related to what they could “actively do”. Posting content, pictures, or videos online and giving an e-mail address, phone number, or credit card were clear harms that participants reported to know. Several participants explained not feeling influenced by TikTok because they did not post any content. Similarly, some participants declared being very cautious when platforms requested for personal data or declared to read what is deemed “important”, like credit card statements or financial information. For instance, debriefing P10's user test, in which she had accepted

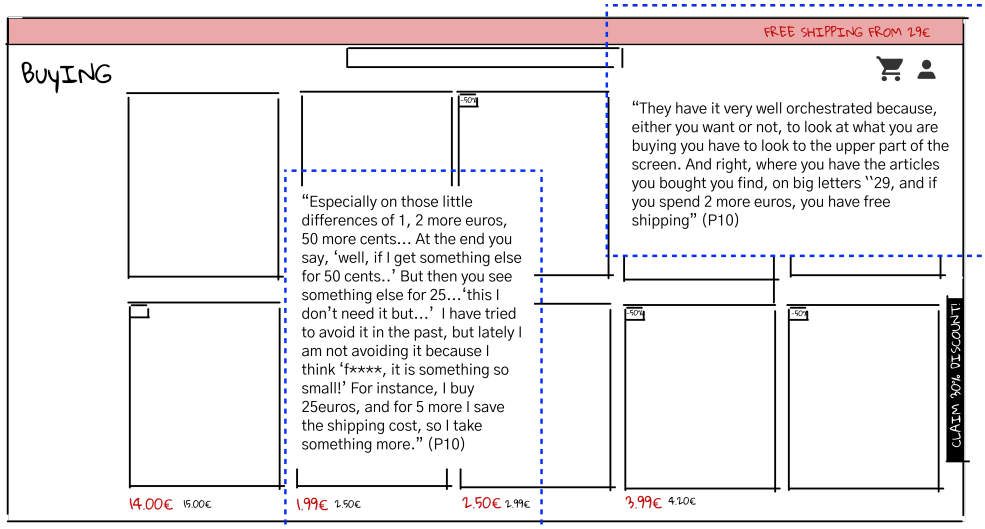


Figure 9.7: Sketch of the platform participant P10 describes during the interview.

the cookie consent banners, she exemplified the binomial between tangible and intangible privacy harms.

P10: *Yeah, it's just I always accept everything. I mean, I know it is very bad, but it's just reading everything is like...uff, I say 'well...'. The typical things from Google, cookies and such, I accept them. Or things that pop-up in webpages of whatever, so... I don't know. I only read what it is like important, important. From there I just say 'yes' to everything because it is very long. There are questions that are very long.*

Interviewer: And what is the 'important, important'?

P10: *Well, it depends on the webpage. For instance, on social networks, there are some parts that are important, for instance, about your data or what you can share. That information I read, because it can be harmful. On e-commerce webpages or so, I always try to pay attention that the data is not registered; for instance, when they say something like 'we are using your data for future purchases' or things related to credit cards, indeed I do read it just in case because I have already experienced some problems."*

P10 normally accepts everything when she perceives no potential harm for her. Therefore, she pays a lot of attention to platforms sharing content or introducing financial information because it might cause her harm. Similarly, P18 explained that only when asked for something like "bank account data", she stopped filling a form without assessing the sensitivity of other information. When P1 sees something related to money, she does not follow-up the e-mails, but she has already given up some information.

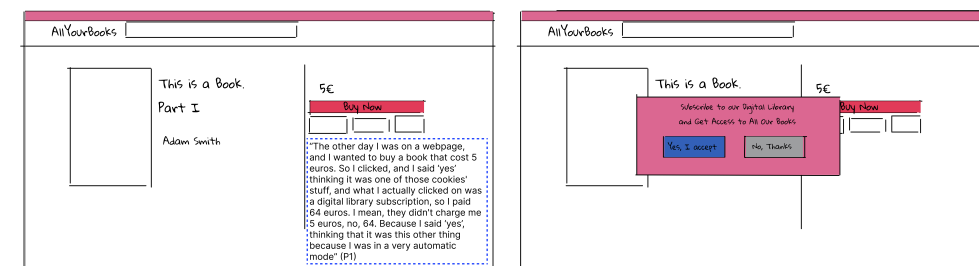


Figure 9.8: Sketches of the interface that participant described. This interface was not shown to the researcher but composed with the information participant described.

Similarly, participants associated cookie banners with ads because that is what they directly perceive. In general, the participants who seemed more aware of cookie banners functioning related it to the advertisement. P15, for instance, did not want to see recommendations on ads, so she would not accept the cookie banners. Noteworthy, during this interview, P15 had accepted all the cookie banners provided to her.

P15: *There, as it is the flights' website, and it will depend on, for instance, it is trustworthy or not, because I usually don't accept cookies, I don't like it, so...*

Interviewer: Why? Why don't you like it?

P15: *Because...I mean, I don't know. I feel it has to do with the fact they are sending me stuff...or that later on [ads] are popping around. I don't understand the topic very well, the term, but I usually don't accept unless on the website they ask you to accept them or that you accept the mandatory ones, the ones that are necessary. If necessary, I go to the settings, except the ones they ask me to accept [...]*

Interviewer: And why did you decide to accept this one?

P15: *Because...I mean, it is a webpage, in theory, that is trustworthy for me. So I accept cookies, I don't see the issue of accepting them.*

Cognitive burden

This type of harm was found in different conversations with participants. While commonly, they did not identify “cognitive burden” as harm itself. Indeed, I should note that it is a very specific academic term to organically bring up during a conversation; there was a trend of finding annoying, repetitive and intrusive certain manipulative patterns. Although they had normalised their existence and felt they could control them, within their experiences, they reported a continuous fight against these patterns in multiple scenarios. When P14 discussed his impressions of some manipulative patterns, he explained how he understands *“it is normal from the point of view of the company, but as a customer, it's a pain”*. This cognitive burden is what triggered on so many occasions that participants would not pay attention to what they were accepting; thus, it is *“unthinkable that any user would read such things”*, reported P1 when referring to consent banners.

These techniques were more problematic in some contexts, like when users were in a rush or when they simply felt the burden was so high that it was insurmountable. Similarly, it becomes problematic when participants are forced to stick to a specific service — i.e., in the case of websites with pirated content, which was often reported. However, this cognitive burden was present not only in pop-ups and nagging techniques but also in other attentional deceptive patterns. Thus, participants would report this excessive cognitive burden when they receive notifications, or even when they see videos on TikTok. On such video platforms, they would struggle to recognise when something is an ad because sometimes it is not visible. Thus, participants reported video-disguised ads on social media platforms as concrete deceptive instances. This is not only presented as a cognitive burden but also makes them spend more time on content they would not want to see. When asked P10 if she had seen disguised ads such as the one provided, her experience encapsulated well the role of the interface in these disguised ads:

P10: “On TikTok. Because there are a lot of people that advertise makeup, or a book, or whatever, and they sell it as a normal video, and you have seen an ad without notice. L’Oréal and these makeup brands that collaborate with tiktokers, so the tiktokers give you a video, and then you read the description, and it says down there “advertisement”: and you swallow an ad without notice. Those are the ones I always end up falling for without noticing.”

Content-related harms

Participants reported several harms related to content rather than the interface. Misinformation caused by hoaxes, people, and influencers was a common threat identified by participants. Similarly, identity-related harms were also commonly mentioned. Participants see how young people are influenced and feel like losing their identity: “everyone is connected and does the same things”. Concerning identity, one participant received content about gangs and was afraid of that type of content spreading among youth. She usually scrolls it down very quickly. Seeing such content may be related to her geographic location, connections, and profile. This is another example of how the profiling of the participants can fuel harm in different ways.

9.4.3 Resistance Mechanisms for Online Manipulation as a Response to Harm

This theme encapsulates the strategies participants mentioned to resist manipulative designs. There was a recurring pattern here: having already experienced harm triggered participants to develop variegated resistance mechanisms. P1 and P10, for instance, would try to pay as much attention as possible because they suffered from subscriptions (P1) or by giving their data (P10). The harm would put them on alert of something “undesired” happening and make them combat it, being the narrative of using platforms as a “fight” or a very pervasive burden. P12 described how she “*had and internal fight with [Instagram]*”.

The reaction to harm is important, especially for less tangible or visible harms. For instance, we reported above that P1 was tricked into giving her data to a website that later asked for money, and she directly closed the page. However, she had already provided data. While she felt manipulated and dealt with it by disengaging, she did not exercise right over her personal data, perpetuating the damage.

The resistance mechanisms are diverse and correspond to different stages of the interaction. However, they seem related to the idea of self-efficacy participants have about the landscape of manipulation and online harms; hence, participants develop the strategies according to their own perceived 'capabilities'.

Disengaging as a coping mechanism

Several participants declared they would disengage from using a platform, website, or service after experiencing harm. P6 explained how, after suffering a scam, she decided not to buy online anymore. After experiencing harm online with a deception technique that led to malware, P3 does not use the computer, an outcome for P2 after her incident as well. When P2 tried to download some content, she experienced adware in the form of multiple nagging pop-ups until she decided to turn off the computer. Since then, she prefers to use the phone.

Another common way of disengaging from the platform is "hiding it". This technique was common in conversations about social networks. Participants described different strategies to avoid these platforms harming them: deactivating notifications and deceptive attention patterns (P1), having detox periods (P12), hiding the phone (P6) using a family member, or distract them (P17). P12 explained that feeling of anxiety or obligation every time she receives a notification. She wants to feel she has agency to choose, rather than the application controlling her; therefore, she prefers to deactivate — when not uninstall — the application for her well-being sake.

Information and Understanding is Vital

Participants expressed the importance of information and understanding. On the one hand, after experiencing some harm, participants would pay so much attention to the information and try to read everything as much as possible. Indeed, more information is what participants claimed to be missing when exposed to manipulative designs: there was a common belief that the more information presented, the better to avoid manipulation. However, there is also a common belief of seeing themselves as incapable of understanding that information. For instance, P6 was always very cautious and fearful, therefore she would always read everything very carefully, coming from a feeling of distrust. However, she would sometimes not understand and have to ask her partner for help because "*she believes everything*".

P7, who is not very fluent in the language, accepts the consent banners because he thinks that is going to help protecting his privacy. He had received some talks about it, so it rang a bell for him, but he did not have a full understanding of it. This was a common trend even for participants who were fluent in the language: it was unclear to them why they were rejecting

or accepting cookies. *"This is something about the ads, right?"* asked P3. She explained she rejected cookie banners because *"she did not feel comfortable giving access to [Google] to her images and sound"*, although the banner that popped up had no reference to such access. Noteworthy, in the interactive tasks probe, none of the participants that selected one of the highlighted options because of the price, understood they were selecting a more expensive option. Participants who confirmed they did not read the consent banners expressed a similar lack of understanding, while others declared they did not do it because they were used to them or simply did not care.

Mechanisms of Resilience: Leveraging Resources to Resist Manipulation

In this category, participants, aware of their harm, try to respond to it with the resources they have. They leverage these resources to try to recover from the harm. Hence, "resilience" seems an appropriate term for these mechanisms. These mechanisms are divided into the following categories: *seeking responsibility/accountability*, *asking for help*, and *re-appropriating the platforms*.

Participants would *seek accountability* from those they believe have the power to do something; here the platforms. They would open complaints and report problems to social media, where most of their experiences of harm took place. Participants also ask external actors for help, like banks or similar entities that seem at their disposal and seem to be accountable. P15 reported a problem on Instagram, but *"it sometimes depends on 'the Internet' "*, meaning that the platform has actual control over what happens. She complements the example with the most famous marketplace platform in Latin America, Mercado Libre. P15 says the platform has the power to protect consumers when harm happens, but it is out of users' control that any action is taken.

P15: *'if Instagram does not take action, nothing is going to happen, same than the other webpages [...] If you see that [someone] has scammed you, and you write a comment about it, they can remove it, and the page... like Mercado Libre, again, does not react and does not eliminate the person [that scams you].*

Some participants also refer to *asking for help* when they face a potentially manipulative experience or already experience harm. For instance, P6 would ask her boyfriend when doubting online. Similarly, P12, P17 or P18 would quickly ask friends and external parties to face the situation.

Participants decided to deal with the harms by *re-appropriating the platforms*, and using them in creative ways. When she was tricked into a subscription to a book platform for a higher amount than promised, P1 decided to download as much as possible and explored the option of re-selling them. Similarly, P18, to avoid the harm coming from the virality of social media uses the application to make personal videos that she does not share. Those videos serve for self-reflection only. Trying to avoid similar harm, P19 has two accounts, a private one and a public one, so she does not share excessively. In another context, when P6 started to receive calls because she was tricked into giving her phone number, she "responded as if she were her

mum” so the providers lost interest in talking to her. A dissonant note in the group was P16. Although he was the only participant who reported to use a technical measure to avoid being located by service providers with a VPN, it still shows the idea of leveraging own resources since, given his conversation and reported data, seemed to be one of the most digitally skilled participants. While these techniques do not necessarily resolve all participants’ problems with felt manipulation, they represent well the non-passivity of users, who do whatever it is in their hands to resist manipulation and its associated harms.

9.5 Discussion

In this section, I discuss what the HCI, CSCW and design communities can learn from the experience of manipulation to rethink countermeasures that account for vulnerability. I, therefore, dive into what these experiences with manipulative designs mean for populations in situations of vulnerability. How users interact and “resist” these manipulative techniques relates to their imaginaries of manipulative designs. First, I explain how the idea of normalisation in these designs relates to the users’ self-efficacy of protection and why it is fundamental for manipulative design scholarship. Second, I explain how such normalisation leaves users to believe that information will help them to resist, as well as the different consequences. Last, I discuss how these experiences can relate to digital inequalities.

9.5.1 The Normalisation of Manipulative Designs

Participants’ resistance strategies are intimately related to their perception of manipulative designs and the feeling of normalisation of these designs. Participants expressed normalisation towards a system that tries to steer their decisions. The rationale behind this *status quo* is that every service provider has the right to sell. Therefore, manipulative designs are marketing strategies for users, which they have to dodge or be responsible for not doing so. While at a macro level, users are used to these strategies, within the micro-interactions that the system allows, they are ultimately responsible. Thus, they would not blame the system they are trapped in, but themselves, because they are presented with a tool not to fall for them: rejecting the options, closing the banners, or closing the applications. There is an idea of resignation; they live under a status quo in which others could change things if they want — e.g. platforms —; hence, users feel they are responsible for resisting.

These results show the relationship between users’ perception of manipulative designs — and the system they are embedded in — and their self-efficacy to resist them — how capable they see themselves of protecting themselves from these practices. The idea of normalisation resonates with previous literature on manipulative designs, mainly in the context of cookie consent banners. Habib et al. [184] show how 50 percent of the participants that accepted cookies did it because they knew that was the way to continue to the service. Utz et al. [442] also see people accepting because they believe they cannot accessed otherwise. Similarly, Nouwens et al. [336] show how participants who have accepted everything did it to make

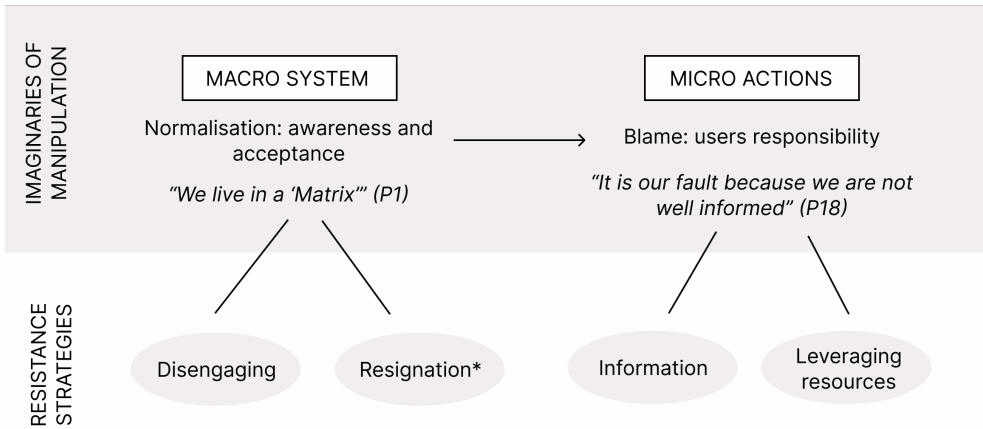


Figure 9.9: Model of the participants' imaginaries of manipulation and their corresponding resistance strategies.

the banner disappear and continue with the service. However, a notable difference between those studies and the present results is that a great number of participants who rejected cookie consent banners showed concerns about the system behind the data economy and consequent privacy concerns — so consciously acting about the decision of how they rejected cookie banners. Our participants were not able to discern those aspects of the data economy.

At this point, it is essential to note the participants' age, since participants had their “digital socialisation” during a time period in which user experience was already dominant on the websites, and manipulative designs already flourished on the web; especially cookie consent banners since 2001 [222]. It seems, therefore, logical that they are normalising these designs in cookie consent banners. If, on top of that, they do not have an environment to learn from and be supported, to realise there is another way, it is complicated for participants to imitate that behaviour — what is called vicarious experiences in psychology. Indeed, van Dijk and van Deursen [446] explain how digital skills are commonly learned via informal education by asking others for help and seeing them in action. Therefore, it is a challenge for researchers, companies and policymakers to help users exit that state of habituation to prevent them from manipulating designs harms.

This normalisation also carries the idea that perceiving manipulative designs and recognising them as problematic is not the same for users. These results seem to be a practical manifestation of the experience of manipulative designs as an invisible phenomenon theorised in [381] and that resonates with previous studies that aim to quantify reasons why users identify manipulative designs. M. Bhoot et al. [286] relate the level of frustration, and the frequency of occurrence showed to be a predictor for manipulative designs identification [286]. In M. Bhoot et al. [286], participants declare they are not tricked, yet unable to identify ma-

nipulative designs, resonating with Bongard-Blanchy et al. [45] and Di Geronimo et al. [111]. This idea is crucial for the research community on manipulative designs. Future research on manipulative designs should take it into account when it comes to making methodological decisions to study the experiences and behaviours of users in the presence of manipulative designs.

9.5.2 Information and Resistance: The Knowledge Paradox

One of the strategies that participants report to prevent them from falling into manipulative designs is using information to make their decisions. As part of that responsibility they believe to have, participants feel that if they read more or better, or if the application provides more information — or information they can understand —, they could resist manipulative designs. This is not the first time we see the idea of gaining agency and self-efficacy through the use of information. Testing different interventions, Schäfer et al. [388] quantifies the feeling of agency shown by participants. In their study, participants are afraid of coercion as a solution for manipulative designs: they do not want any imposed solution, expressing this need for agency. This resonates with the participants in this study, who demand the need to reject buttons and interface elements that give them control over their interactions. With that underlying assumption, some studies argue for increasing the knowledge of users. Tjostheim et al. [434] have tried to use board games to increase knowledge on manipulative design, which proved not effective, and Chen et al. [80] have designed an intervention to provide more information about the manipulative design during the interaction, which proved to be effective for reducing consent on cookie banners.

Building on theories of protection, like protection motivation theory, there is an assumption that if users' knowledge about manipulation increases, the appraisal would also increase, triggering protection for users. With this rational Strycharz et al. [427], show how more information did not affect the user's decisions when opting out of personalization. Klütsch et al. [237] did not find a relationship between providing more knowledge within the interaction and cookie acceptance in the presence of manipulative designs. Indeed, although this study does not claim any generalisability, the results provided showed that none of the participants understood they were selecting a more expensive option when selecting a flight, which is in line with existing literature. Berens et al. [31] shows in a controlled experiment on cookie consent banners how the information provided in the banner did not influence its acceptance, resonating with previous literature on the topic [250, 251]. Participants show paradoxical behaviour; they demand more information that later on does not seem to be relevant to the interaction.

This resonates with several studies that have analysed the different barriers on users' journeys for newspaper subscription cancellation [401], social network cancellation [390], or e-commerce websites [316]. Participants experience damage and defend against it while accepting and normalising the existence of these mechanisms because they believe they can resist them. Looking at theories of motivation [107], it seems that users have a sense of competence over the interaction when they face a manipulative design in which trust plays an im-

portant role. When participants say they trust instances with manipulative designs and that they would not feel prey, it is because those specific providers manage to fulfil those needs until the harm comes, but it is a feeling of false agency. While some papers [399] have theorised about the feeling of false agency, this is the first time empirically coming from users' data. In this way, by understanding users' imaginaries of manipulation and consequent coping strategies and their associated challenges around users' agency, I present a first approximation of what resistance means for users in situations of vulnerability.

With this discussion, it is not my intention to blame and shame the user, falling into victim-blaming positions that only harm the most vulnerable collectives. On the contrary, I aim to call to action to researchers and policymakers when they position the efforts on the user by asking for more digital literacy and more knowledge on manipulative designs. I want to invite the community to rethink what is actually what we should make users aware about: the logic of the data economy system, rather than the mere recognition of manipulative designs. So this raises the question: how can the HCI, CSCW and design community contribute to that end within their interventions?

9.5.3 The Relationship Between Manipulation Resistance and Inequalities

This normalisation might be more problematic for users in situations of vulnerability, who might get used to them while having to make other trade-offs. In Sánchez Chamorro et al. [382], teenagers had to make trade-offs within the interface: time, money, and privacy. In this study, similar contextual and personal conditions mediate the harm. Not only is the strictly economic aspect important, but the lack of support and loneliness also have socio-economic relationships. Looking at socio-digital inequalities theories, social support is related to socio-economic conditions [446]. This means that some populations are put an extra burden on resisting such harms, as these designs are more likely to cause damage.

This relationship with the socio-economic position is noticeable in developing their coping mechanisms. Coping by disengaging is a consequence of not feeling capable of conducting a task or not being able to recover from harm, as it is shown in the results. Perceiving themselves with a lack of skills and not having proper support will turn into users with more vulnerable positions, being excluded from the digital system, which, ultimately, may contribute to reinforcing these inequalities [198, 196, 195]. Similarly, the idea of resilience mechanisms as a coping strategy comes again from the user's perception of self-efficacy to resist. Users leverage their resources in the ways they seem capable of doing so. This is important for the HCI and design community because it is a potential entry point to design interventions taking vulnerability into account. How can we help users to leverage their resources so they feel empowered to resist manipulation? This idea seems crucial for the community that often talks about empowering users to face manipulative designs, but it is unclear in which ways that can happen. Leveraging user's strengths and resources is a starting point for designing interventions that increase resilience and empower users in situations of vulnerability.

9.5.4 Limitations and Future Work

This study presents some limitations in its design. Given the nature of the study, it was expected that participants would bring usability issues during the interview, which was a limitation for the subsequent data analysis. I tried to overcome such limitations by laddering different critical incidents. While the initial introductory question was meant to elicit incidents during participants' online interactions, the interview guide would ladder into moments in which users might have felt deceived and, ultimately, manipulated.

While data triangulation was conducted to provide a variety of entry points to understand situated experiences with manipulative designs, the fact that participants were shown static images rather than an immersive website or application limits the data provenance. Similarly, the fact that the interactive tasks and the probes were conducted and shown in the desktop versions instead of the mobile phone ones can also limit data provenance since participants were generally more acquainted with the mobile platforms than computer versions, which they did not use often. In turn, this offers new opportunities for research that take into account the differences in experiences in a multi-modal approach and with different devices. A limited set of literature explores manipulative instances in mobile applications [111, 180], and some preliminary research indicate differences in behaviours in presence of manipulative designs depending on the device [447, 34].

I designed the study under the assumption that the more familiar participants are with the website, the easier to circumvent the manipulative designs are. Therefore, having diversity in familiarity would allow me to see the cases in which participants have adapted versus those cases in which they did not have time to adapt to unfamiliar websites. It is not the intention of the study to make any experimental setting, so there was no quantification of such phenomenon, but to see in their own experiences how manipulative designs could work regardless of familiarity. Therefore, the extent to which participants were acquainted with the applications provided in both the interactive tasks and the interview with probes could limit the results — i.e. participants did not often buy flight tickets and felt limited in the interaction such platform. I tried to overcome this limitation by choosing two tasks in which one scenario would be familiar to them. Indeed, the fast fashion website was very popular among all the participants who reported using it often during the interviews.

9.6 Conclusion

This study aimed to understand the experiences of users with low levels of digital skills with manipulative designs. By understanding their perceptions and imaginaries of manipulation and coping strategies, I have outlined the problems that the normalisation of these practices brings. This paper sheds light on the problem of recognition in opposition to the perception of manipulative designs and what that means for the set of actors that try to fight these designs. By reflecting on these experiences, this paper deepens into the idea of agency, resistance and self-efficacy among low digital skill users. It, therefore, helps to unveil what resistance means

for users in situations of vulnerability and opens spaces for new perspectives in the study of manipulative designs.

9.7 Chapter Takeaways

By looking at the experiences of young adults with low digital skills, this chapter provides a more nuanced understanding to the problem of recognition vs perception of manipulative designs and what that means for the set of actors that try to fight these designs. Similarly, it also deepens into the idea of agency, resistance and self-efficacy to the system of manipulation and concrete manipulative instances, and reflects what that means for the different actors. Furthermore, it explains the imaginaries of resistance of users, which can be leveraged by designers to create counter-measures for manipulative designs.

Part IV

Design Spaces to Mitigate Vulnerability to Manipulative Designs



This part discusses the implications of looking at the users' experiences with manipulative designs through the lens of vulnerability.

Chapter 10 offers an exploration of user experiences with manipulative design in a temporal way in order to elicit drivers of vulnerability through which look at intervention spaces. With an overview of the current intervention spaces, it offers a reflection to rethinking existing intervention spaces and including new ones that leverage vulnerability to protect users with their own resources.

Lastly, **Chapter 11** offers a general discussion with a focus on how experience supports the understanding of vulnerability drivers to manipulative designs and their implications for the design community that aims to protect users from manipulative designs.

The dissertation ends with a call to action for different stakeholders and the research community in manipulative design to broaden the understanding, methodologies and lived experiences of users within the interaction with manipulative practices.

Mapping the Experience of Online Manipulation: Rethinking Resistance Strategies Against Manipulative Design

This chapter is based on the following publication:

Lorena Sánchez Chamorro, Romain Toeboesch, and Carine Lallemand. 2024. Mapping the Experience of Online Manipulation: Rethinking Intervention Spaces and Resistance Strategies Against Manipulative Design. 2024. Submitted for journal publication

Abstract. The pervasiveness of manipulative designs (so-called dark patterns) is associated with diverse types of harm to the users. While scholars increasingly focus on mitigation strategies, a better understanding of the experience of online manipulation and where the conditions for vulnerability occur is essential to identify new intervention spaces against manipulative designs. In this paper, we first present a rapid literature review to understand the current landscape of interventions against manipulative designs. In a second study, we map the experience of online manipulation based on two qualitative studies with populations at risk of vulnerability. Our findings depict the temporal, contextual, and ecological factors that mediate the relationship between users and manipulative designs, and identify drivers of vulnerability that can inspire interventions. Based on our journey map of online manipulation, we ideate and discuss new intervention spaces to resist manipulative designs or mitigate their harm.

10.1 Introduction

Manipulative designs¹ — so-called dark patterns — are “instances where design choices subvert, impair, or distort the ability of a user to make autonomous and informed choices in relation to digital systems regardless of the designer’s intent” [174]. Due to their associated harms on users, manipulative designs are raising attention among scholars and policymakers [341]. The extent to which these elements lead users to attentional, privacy, emotional, or economic harms [341, 382, 181] — especially among specific users in situations of vulnerability [341, 340] — has called the design community to work around the design of counter-interventions supporting users in coping with these manipulative instances and preventing their harms [390].

To better understand the problem of manipulative designs, the HCI community has intensified research on its effects on users [171, 76], investigating multiple contexts such as social media [390, 232, 309, 228], streaming platforms [78, 119], conversational user interfaces [307, 347] or cookie consent banners [31, 46, 38]. A vast part of this increasing body of literature is looking at changes in behaviour and perceptions caused by manipulative designs, e.g., whether a User Interface (UI) elicits certain behaviours or not [38, 31, 45, 286, 76]. Despite their contributions to the understanding of the workings of manipulative designs, these approaches still face translational challenges and have limitations in ecological validity, as they do not take the contextuality of user experiences [38, 31, 439, 381] into account.

Similarly, the research community is concerned about the harms associated with manipulative designs, and several studies discuss potential solutions to address these [389] in the form of UI design interventions [31, 162], behaviour change strategies [281], educational interventions for users [45], training for designers [83], and technical solutions [80]. However, HCI studies mainly investigated or discussed these ideas in decontextualised ways. Many suggested solutions do not yet consider temporality or the dynamic between different users and their environment.

We argue that the conception and design of counter-measures to manipulative designs would benefit from a more contextualised understanding, where the temporality of the interaction and the user’s environment are taken into account. To reduce users’ vulnerability — understood as a position in which users are more likely to be impacted by specific harm or less likely to recover from it [289] — to the harms of manipulative designs, we need an understanding of the conditions under which users are more vulnerable. These conditions, so-called “drivers of vulnerability” [289], are dependant on context and therefore difficult to uncover and mitigate without approaches that help understand the situated and contextual experience of manipulation [381]. To do this, it is important to acknowledge the relational nature of the experience of dark patterns [382, 383]: the long-lasting effects of manipulative designs and associated experiences make it so users only perceive them when they are

¹The research community is studying this phenomenon using a variety of labels, including deceptive design, nudges, anti-patterns, and most dominantly, “dark patterns.” Following the ACM recommendations on diversity and inclusion [143], we opt for the use of manipulative designs as a term that both more inclusive and more accurately describes the phenomenon

impacted. Adopting a temporal lens when looking at users' experiences with manipulative designs would hence favour understanding the drivers of vulnerability to these designs, and in turn open more opportunities for the design of counter-measures.

In this paper, we adopt a temporal perspective to the experience with manipulative designs [170] to elicit drivers of vulnerability to manipulative designs. This will ultimately allow us to open solution spaces to manipulative designs that consider vulnerability. As a first stage, we conduct a systematic review of the main interventions to counteract the effects of manipulative design on the user side that has been discussed in the literature. As a second stage, building on previous work highlighting the benefits of experience mapping as a research synthesis tool [213, 266, 457], we create an experience map of manipulation as a first step to understanding the temporal aspect of manipulative designs from the user's perspective. To do so, we combine data from two studies in which users in situations of vulnerability talked about their experiences of manipulation, by paying attention to the experience along the different stages of the interaction, together with the people, technologies, and contexts that play part in this. We finally discuss the user experience map as a tool to derive interventions that account for vulnerability and put it together with the existing ones to give an overview of the different solution spaces, strategies and interventions that can be used to counteract manipulative designs harm.

Through this paper, we contribute to HCI research and practice on manipulative designs by (i) providing insights into the contextual and experiential drivers of vulnerability to manipulative designs, derived from empirical data, (ii) systematically mapping existing interventions against manipulative designs, (iii) opening opportunities to further investigate new solution spaces against manipulative designs, particularly suited to populations at increased risk of vulnerability. We hope these insights will aid the HCI and Design communities in the development of a more diverse set of solutions that help reduce users' vulnerability to manipulative designs.

10.2 Related Work

10.2.1 Experiences of Manipulative Designs

Prior work has examined the effects of manipulative designs on users, from a behavioural perspective — looking at the effects of design elements on observable behaviours [162, 38, 31] — and from an experience-perceptual approach — investigating users' perceptions and experiences [45, 286, 163]. From both approaches, one can elicit key points contributing to experiences with manipulative designs.

Awareness of manipulative designs

Several studies investigated the awareness of manipulative designs among users, showing that users present difficulties in identifying what what researchers categorise as manipulative designs. Di Geronimo et al. [111] and Mildner et al. [308] explained the blindness of users to

manipulative designs. Similarly, when shown manipulative designs, Bongard-Blanchy et al.'s [45] participants were uncertain whether the designs might influence them or not, although there was general acceptability towards them. People were aware that services might influence them in general, although they feel it is more likely that others are influenced rather than themselves. With this rationale, Gray et al. [163] explored the experiences of manipulative designs as “felt manipulation on users”. In their survey, participants generally reported perceiving manipulation when being asked about privacy, felt threatened about security, or asked to pay for products.

Some studies have tried to explain the reasons behind this ability (or lack of) to identify manipulative designs. There is an assumption that awareness and identification of manipulative designs are related to the use of technology. Bhoot et al. [286] found a positive association between the frequency of manipulative designs and perception of trustworthiness. However, using the Affinity Towards Technological Interaction Scale [148], Voigt et al. [453] did not confirm any association between participants' affinity for technology and recognition of manipulative designs.

Few studies have looked at how socio-demographic variables affect users' relationship with manipulative designs or their experience. Bongard-Blanchy et al. [45] and Luguri and Strahilevitz [280] showed how levels of education impacted the recognition of manipulative designs and resistance. Similarly, Bongard-Blanchy [45] showed that older participants had more trouble finding manipulative designs, resonating with Avolicino's [15]² findings, who found older adults to be less acquainted with the service. Sánchez Chamorro et al. [383] identified a common lack of self-efficacy to resist manipulation in participating older adults, who did not — nor wished to — identify instances of manipulative designs. The impact on children and youth is a concern among policymakers OECD because they are more prone to take risks [340] and less resilient. Radesky et al. [367] explained the risks related to the high presence of manipulative designs in mobile applications for kids is further exacerbated for those with lower socio-economic status.

Perceptions of Resisting Manipulative Designs

The capacity that users perceive to resist manipulative designs — namely, self-efficacy — has been explored in some studies. While there is no unified definition of ‘resistance’, scholars commonly consider this to be when users avoid falling for the manipulative design — regardless of whether they actively seek this avoidance or not. Bongard-Blanchy et al. [45] observed a relationship between people's perceived self-efficacy to resist manipulative designs and their capacity to recognise them, resonating with Maier and Harr's [287] perceptions on manipulative instances. When shown manipulative designs, participants assessed their impact, but also “accepted them” because they find them easy to elude “by clicking the X”.

The role of users' perceived agency has gained importance in studies on manipulative designs. By investigating possible countermeasures, Schäfer et al. [389] discussed the dichotomy

²We believe that Avolicino's results in relation to age must be taken into account cautiously since their oldest cohort is “over 35 years old”.

between the user's desire to remove confusing information, while being in control of what is removed, thus demanding agency. In this line, Lukoff et al. [281] explored how to help users regain agency when facing deceptive attention capture patterns. Chaudhary et al. [78] and Lyngs et al. [285] made similar claims for giving users' agency when they face manipulative designs.

Some studies have explored the assumption that with more knowledge, users will be more likely to avoid the effects of manipulative designs during the interaction. Chen et al. [80] thus developed a tool that provides knowledge to users within the interaction. Through the evaluation of the tool with users, it proved useful to help users recognise manipulative designs. With the same rationale, Tjostheim et al. [434] created a board game aimed at increasing knowledge of these instances among teenagers, although it did not prove useful for such purpose when it was tested. Hence, it is still unclear from these research works, whether more knowledge about manipulative designs can help users. Additionally, Strycharz et al. [427] found that technical knowledge in consent notices do not impact opt-out for personalized ads. Testing this hypothesis on manipulative designs concretely, Klütsch et al. [237] did not find a relationship between providing knowledge to the user when the interaction takes place and user consent.

The Role of Trust in the Experience with Manipulative Designs

Trust in the service is a recurrent underlying topic in the different studies on manipulative designs. Gray et al. [163] explained how trust in the service and the different elements that contribute to distrust in users are fundamental triggers of the experience of manipulation. Bhoot et al. [286] found a positive association between trust and “the physical appearance of the platform” and trust. Tuncer et al. [439] explore “perceived benevolence” as a sub-dimension of trust and showed a negative relationship between the presence of scarcity cues and the perceived benevolence of the services. Similarly, Mildner [306] explained participants need trust in the development of ethical conversational user interfaces; however, this development of trust can flip since, as Sanchez Chamorro et al. [379] explain, UX/UI designers want to build trust on users to influence their behaviours.

Emotions and psychological harm

Gray et al. [163] showed how the experience of manipulation came associated with negative experiences and emotions — i.e distress, upset, guilt, fear, hostility, irritability or shame, among others —, in line with the results of Avolicino et al. [15]. Bhoot et al. [286] found that frustration correlated with the identification of manipulative designs. In their study with teenagers, Sánchez Chamorro et al. [382] also found the trigger of insecurities and impulsivity in users was also associated with manipulative designs. Similar negative emotions of impulsivity were found in Chaudhary [78] and discussed by Moser et al. [316]. Chaudhary also explains how participants show negative emotions and regret when they realise they have continued watching unintentionally. In the same line, the study of Tuncer et al. [439], on scarcity cues

showed a relationship between negative emotions and the presence of these design elements.

10.2.2 Drivers of Vulnerability to Manipulative Designs

Vulnerability is a multifaceted concept that has been discussed in different domains [290, 288, 138, 283], including in manipulative design research, as a greater susceptibility to harm [375, 340]. Vulnerability has been traditionally discussed from legal theory perspectives [283, 288, 138], where it is understood as layered, relational, contextual and situated [283, 290, 288]: while all users are vulnerable, some more than others. Applied to the online domain, Malgieri [288] explains how for vulnerability to happen, elements have to fuel this susceptibility to harm: drivers of vulnerability. This relationship between vulnerability and manipulative designs is raising concern among scholars and policymakers [341, 340, 382, 472, 383]. Indeed, the OECD [341] has reported some drivers of vulnerability related to the market, and to personal conditions of users, focusing on their (lack) of rationality. Yet, they do not consider the contextuality of the user experiences, making it an underexplored topic to date.

To study vulnerability to manipulative designs, we adopt the view of manipulative designs as contextual, situated, and relational [382, 383]: users can only perceive them when they receive an impact. This aligns with Gray et al. [169]'s theorisation of the experience of manipulative design as a mismatch between users' expectations and the interaction feedback. For similar reasons, Gray et al. [163] used the experience of manipulation as a proxy for the effects of manipulative designs. They extended the ideas of users' assessment of manipulative designs from Maier and Harr [287] and set a preliminary timeline of the temporality of felt manipulation: initial judgment, an inspection of details, felt persuasion, general conclusion, undesired interaction and negative results from interaction. Gray et al. [163] explained how since the very beginning of the interaction, users assess whether they can trust the interface or not, reaching to a point in which they perceive unwanted outcomes.

With this work, we aim to shed light on specific pain points that drive vulnerability, and their corresponding design spaces. By looking at the temporal experiences with manipulative designs of populations in situations of vulnerability, we contribute to this gap.

10.2.3 Interventions to Counteract Manipulative designs

The research community has proposed different types of interventions to counteract manipulative designs. A few of these have been tested through experimental studies, whereas others are only mentioned by authors and do not yet have empirical support. The level of granularity also differs, from precise interventions that can be operationalised in a specific context to more generic calls to action (e.g., "we need more regulations"). The OECD [341], aware that users fall for manipulative designs because of the existence of biases and shortcuts, advocated for different comprehensive interventions also supported by scholars [388]: regulatory interventions [181, 347, 280], educational-related interventions [111, 247, 287], or the creation of automated tools [111]. Bongard-Blanchy et al. [45] discussed a potential matrix of along educational, design, technical, and regulatory domains, with different purposes: elimi-

nating, resisting, detecting and increasing awareness of manipulative designs. Other scholars have called for social-related interventions and challenges [458, 47, 383]. However, it remains unclear what different stakeholders can do, and what their boundaries are among the different solution spaces. Mapping the existing interventions would help to gain some clarity, not only on the possibilities but also on the potential challenges and limitations.

The OECD [340], recognising the difficulties of studying online vulnerability, made a call for finding new methods that can look at vulnerability in a contextual and situated way. Yet, while the HCI and design research community works around design outcomes, user-centred design counter-interventions are limited [389] and those looking at the temporal aspect of manipulative designs are still needed [163, 170]. The highly context-dependent and temporal aspect of the effects of manipulative designs makes it easy to understand the challenges of designing interventions that effectively work. Thus, to rethink design interventions, it is first necessary to understand the experiences of these manipulative designs as a dynamic and long-term process.

10.2.4 Using Journey Maps to Design Counter-interventions for Manipulative Designs

Customer Journey Maps are tools used to understand the relationships between customers and companies through their products and services [Kalbach as cited in Ismirle [213]]. These tools are common in marketing research and have been extended to the domain of design and HCI research as tools to support the understanding of users in their relationships with design artefacts. Thus, they have been used in different contexts like hardware installation [121], human-soundscape relationships [470], interactive museums [294], healthcare [61], railway experiences [258], runner experiences of women [303], and charity donations systems [212].

The use of journey maps focuses on synthesising the patterns of the user experiences [213] along the temporal journey of such experience, which might prove useful in the context of manipulative designs. The need for a more contextual and temporal understanding of the experiences of users with manipulative designs [381, 170] makes user journey maps a suitable tool to put user data in a temporal and relational aspect — as it allows to understand the role that other stakeholders and actors play [205] — of the experiences with manipulative designs. Lastly, it is important to mention that using a practitioner's tool will allow better communication of the existing research insights and translation of them into design practices.

10.3 Research Approach

Building on the aforementioned works, this paper has the overarching goal of investigating the situated experience with manipulative designs, eliciting the pain points in the manipulation experiences as a means to open intervention spaces that take vulnerability into account. Hence, it addresses the following research questions: (RQ1) What are the main interventions suggested by the community to counteract the effects of manipulative design? (RQ2) How can

looking at the temporal and contextual aspects of the experience with manipulative design contribute to creating solutions for manipulative designs?

To do so, we conducted a systematic review of the literature to understand the interventions that are currently suggested by HCI and design scholars. Later, we conducted an experience map to understand how contextual and temporal experiences could offer new opportunities to design interventions against manipulative designs. Both studies will be individually explained in the following sections.

10.3.1 Positionality Statement

To report these studies, it is important to acknowledge our positionality in relation to this work [238], taking our subjectivity as a resource [161, 51]. The study of manipulative designs as a threat to users online is both an academic and a personal interest of the first author. They have strict values about how problematic these designs can be, which definitely impacts the way they see the experience of manipulation and associated interventions. They share language, and socio-economic background with the participants, and therefore could resonate with some of their experiences. Noteworthy, they have already analysed the data used in this study, which gives them a deeper understanding of the data, but also impacts the analysis. The second author is a design researcher. In their work, they challenge technological paternalism and explore ways to empower people in their use of technology. The last author's main expertise lies in design methods and how they can be instrumental in making sense of participants' lived experiences. They have experience in using journey mapping as a research tool and in conducting design space explorations. The second and third author are not as closed to the participants in terms of experiences as the first author, but offered insightful contrasting perspectives and a more analytical approach to interpreting findings and mapping processes.

10.4 Study 1. Systematic Review of Interventions

In this section, we present a rapid review of interventions against manipulative designs, as a sub-type of systematic review and analysis [188, 319]. Understanding the current landscape of interventions used or discussed to counteract manipulative designs harm is key to understand the main aspects of the experience that the community aims to address to counteract manipulative the effect of manipulative designs. The motivation behind the rapid review as a method resides in the need to depict the directions where the field is moving in order so we can provide initial results to the community on a topic that is rapidly changing and evolving [339]. We reviewed the main interventions addressing manipulative designs in the literature by relying on the initial datasets from the systematic literature reviews of Gray et al. [171], and Chang et al. [76], which already extended the former. Considering the data collection time frame, we expanded this dataset with papers published until 1st June 2024. The list of papers included is presented in the Supplementary Material ³.

³bit.ly/supplementary-material-rapid-review

10.4.1 Methodology

Data collection

We report on our search process, inspired by the PRISMA statement [364]. We selected the ACM Digital Library as a core database for HCI and design-related publications. To remain coherent with the two reviews we build on, the search query was also limited to the term “dark patterns”. We initiated the process with 104 entries from Gray et al. [171], 45 from Chang et al. [76], and 104 new ones from our search.

We checked for duplicates among the three entry points with Rayyan⁴ and obtained a total of 232 papers. Noteworthy, we used the pre-screened database from Chang et al. [76] as their filtering could leave aside some relevant work for us. Building on Gray et al. [171] and Chang et al. [76], we did the screening of title and abstract with the following inclusion criteria were:

- The entry is published in English.
- The record mentions “dark patterns” explicitly in the full text
- The record is published in an archival venue, such as conference proceedings, journal, or government technical report.
- The record includes at least one empirical component, such as analysis of users data, documents, or artefacts.

We also used the following exclusion criteria: master/bachelor thesis, the topic of the paper is not related to manipulative design instances that try to steer users into decisions they would not initially make, there is no empirical component, and there is no traceability for the entry, so the article cannot be found in any archival database. With these exclusion criteria in mind, we excluded 99 entries, including some from Gray et al. [171] and Chang et al. [76]’s reviews. To enhance the comprehensiveness of the review, we included the following papers that met the criteria that were not found in such search: [38, 79, 366, 427, 433, 439, 315, 278]. 139 papers were finally considered in the database for reviewing.

10.4.2 Data Analysis

Through a deductive-inductive iterative process, the first author developed a codebook representing the main categories of interventions discussed in the literature to fight manipulative designs. Interventions are understood as any action conducive to protecting users from potential harms caused by manipulative designs, and we initiated our codebook inspired in the main categories that Bongard-Blanchy et al. [45] use in their discussion: educational, design, technical, and regulatory. Throughout our iterative process, our categories evolved from the initial ones.

⁴<https://www.rayyan.ai/>

As part of the process, we have identified the level of granularity of the interventions against manipulative designs. Overall, prior publications do not share a unified language on the types or levels of interventions. We noted that the terms "interventions", "countermeasures", "mitigation strategies", or "solutions" were often used in a broad sense as umbrella terms to designate anything aimed to fight manipulative designs or mitigate harm to users. Our systematic review, therefore, includes an amalgam of design concepts, testable prototypes, overarching design strategies, or solution spaces identified as "interventions". We refer to the overall landscape of intervention opportunities as the *intervention space*, i.e. the range of opportunities within a system where a stakeholder can intervene to reduce the presence and/or effects of manipulative designs. Realising the different levels of granularity allowed us to find more commonalities and differences between different interventions. (See Table 10.1).

Granularity level	Description	Example
Interventions domain	A domain relates to a particular stakeholder that can reduce the effect of manipulative designs (e.g. creation, development, interaction)	Education of Users — with education users can make themselves less vulnerable to manipulative designs
Design strategy and sub-strategies	General approach and means through which the stakeholder that can reduce the effect of manipulative designs	Providing knowledge — by providing users with knowledge they are better equipped to recognise manipulative designs
Design concept	"Set of ideas for a design that usually comprises scenarios, images, mood boards, or text-based documents" [471]	An educational board game — users can learn to recognise dark patterns through an educational board game
Design prototype	"Manifestations of design ideas that concretise and externalize conceptual ideas" [471]	'The Dark Patterns Game' — A prototype of a board game for teenagers to learn about manipulative designs

Table 10.1: Codes related to the different levels of interventions within the intervention space against manipulative designs

10.4.3 Findings

We offer an overview of the design space of interventions against manipulative designs suggested by the literature, structured by the domain they act on. Most of these interventions are discussed in the papers' discussion sections (often suggested as future work) but not evaluated — we indicate in the text those that have been tested.

Regulation and co-regulation

Under this domain, scholars call for governmental and policymaking actions to collaborate with companies and impose obligations concerning manipulative designs [390, 45, 31, 347]; hence, the main targets of these interventions are companies. Most authors suggest interventions in a generic way (with no primary focus on discussing regulations). This domain has three main strategies: *statutory regulation*, *co-regulation and self-regulation*, and *enforcement*.

Several scholars demand more *statutory regulation*, understood as binding rules from national or supranational actors — i.e. governments or international organisations. There are three sub-strategies around statutory regulation: standardising definitions, establishing binding requirements and banning. On a more conceptual level, scholars ask for better definitions of multimodal aspects applicable to different technologies such as CUIs, IoT or XR — [306, 347, 246, 243, 185]. Specific binding requirements for browsers, consent management platforms [336], choices, reminders, and autorenewals [401] are also suggested. Scholars also ask to ban specific types of manipulative designs that are already known [308, 458].

Co-regulation and self-regulation refer to mechanisms that impose a regulatory framework on companies that are adapted to their contexts, and in which companies collaborate to develop such obligations – e.g. codes of conduct and standards — [365, 59, 67]. Thus, two sub-strategies arise: *standards that can be enforced* [183, 315, 243, 476, 379] — i.e. more concretely Aagaard et al. [1], suggest a “dark patterns badge” that companies can use to prove they do not contain manipulative designs — and *codes of conduct that companies can adhere to* [476, 379] — i.e. with design loyalty rules [243] and design recommendations for consent [252, 419].

The third sub-strategy in the regulatory domain is *enforcement* on those platforms that are known for the use of manipulative designs. More concretely, scholars ask for enforcement of companies that do not have privacy-friendly options as a strategy, enforcing at least the known famous platforms [458], fostering the truthfulness of legal disclosures, requiring feature parity, and promoting holistic investigation enforcement [180]. When it comes to enforcement, scholars consider the *use of automatic tools* as a sub-strategy, under the idea that the known attributes of manipulative designs can be automatised, the idea is that automatic tools can help regulators to enforce and control manipulative designs on website⁵. Some prototypes in this regard are the tools from Kollnig et al. [241] and Chen et al. [80], that are expected to be picked up by regulators. According to Curley et al. [103], automation is challenged by the complexity of certain manipulative instances. Kollnig et al. [241] hence suggest community-driven tools. Plugins that remove manipulative designs have also been explored on Facebook [314]. Both Purohit et al. [366] and Lyngs et al. [285] have tested automatic filters that remove manipulative designs on users, proving to be effective to reduce users’ social media consumption. Chen et al. [80] have developed a tool that recognises and explains manipulative designs to users during the interaction. Nouwens et al. [336] suggest technical solutions to cookie consent problems, by obtaining consent through durable means like the browser.

⁵In the case of automatic tools, authors also suggest them to help users in discovering manipulative designs or filtering them out [241, 366], so we consider automatic tools to be multi-domain

Lastly, scholars argue for *using HCI methods to support detection and enforcement* from policymakers [308, 336, 111, 174]. Indeed, Gray et al. [174] argue for using the ontology of manipulative designs as a tool to support regulators.

Educational and awareness interventions on users

The main objective of this domain is to elevate users' capacity for understanding and identifying manipulative designs, which will ultimately help users to increase resistance through knowledge. While users are the main target of this domain, the stakeholders that are expected to educate users remain open. A set of literature suggests interventions meant to provide more information about how manipulative designs work and create awareness. There are two main strategies: *providing knowledge* to users and *promoting collective action*.

Some studies appeal to *providing knowledge to users* as an intervention to create awareness of manipulative practices, meant to "favour users' agency" [45] (p. 771), and suggest interventions to give users information about how manipulative designs work through mediums external to the interaction — i.e., educational programs or public awareness. These studies refer to education in a broad sense [287, 45, 399, 230, 111, 169, 247, 328, 215], putting the emphasis on users need to be "educated" so they can protect themselves against manipulative designs. One sub-strategy is to provide knowledge *inside the interaction*. Few studies include specific concepts about how to do this. Creating awareness tools like plugins [241], using fear responsibly within the interaction [88], or using reviews [16] to indicate the presence of manipulative designs are some examples suggested. At a prototype level, Klütsch et al. [237] tested cookie consent banners informing information about manipulative designs, yet this intervention did not prove effective in these studies.

Another sub-strategy to provide knowledge to users is the use of *outside the interaction* interventions to raise awareness. Different concepts have been suggested in this area; for instance, Keleher et al. [230] suggest general awareness campaigns targeting daily technologies in mainstream media, as well as promoting a tip line for users to receive help through similar media, Westin and Chiasson [458] suggest to increase awareness of FOMO centric-design. At a prototype level, Tjostheim et al. [434] tested "The Dark Patterns Game", a board game to increase the knowledge of manipulative designs among teenagers.

Promoting collective action among users is another strategy in the realm of education and awareness, addressing power imbalances between users and platforms [230]. Scholars suggest three types of strategies in this realm: *shaming and flagging companies* with bad practices, and using some concepts like community-drive applications [241] as an invitation to collective action among users; *coordinating advocacy and policymaking* for collective action against manipulative designs [279, 235]; and using citizen-science approach [279, 235].

Designers and developers measures

This domain refers to interventions to support designers and developers to reduce manipulative designs within the design process. Designers and developers are a fundamental part of

Regulatory Domain			
Intervention Strategy	Substrategy	Concept	Prototype
Enforcement — <i>Imposing obligations to companies and digital service providers</i>	Privacy-friendly options [46]	Enforcing on most famous platforms [38]	-
	Facilitating detection	Using automatic tools [336], Using the ontology of MD Holistic investigation enforcement [46] Truthfulness of legal disclosures [347] Feature parity [347]	Chen et al. [80], Kollnig et al. [241] -
Co-regulation and self-regulation — <i>Establishing collaborative links between companies and regulators</i>	Standardisation [46, 379, 476]	Dark patterns badge [1]	-
	Codes of conducts (self-regulation) [379] Design recommendations [436]	Design loyalty rules [243] For consent [419, 252] For ethical implementation To avoid manipulative designs	- Lapin and Volungevičiūtė [261] and Sánchez Chamorro et al. [379]
Statutory regulation — <i>Establishing obligations for companies and service providers in the design of technology</i>	Standardising definitions	Multimodal regulation and specific regulations: voice interface [347, 306], IoT [243], XR [246, 185]	-
	Binding requirements Banning [307]	For browsers [336] For CPM [336] For clear choices, reminders and renewals [401] Identified manipulative designs [458]	- - - -

Table 10.2: Summary of interventions belonging to the regulatory domain.

these interventions, but who is responsible for supporting them remains open. In this domain, scholars have focused on three strategies that move around: designers' capacity, designers' knowledge, and designers' responsibility.

Interventions that aim to increase designers knowledge focus on supporting designers with knowledge with technical tools for developers [430], templates, APIs, guidelines or specific frameworks adapted to technologies like conversational user interfaces [306, 347, 124] or extended and virtual reality [246]. On the prototype side, Chen et al. [80], Kollnig et al. [241] and Shi et al. [403] have tested prototypes for manipulative design detection. Lapin and Volungevičiūtė [261] and Sánchez Chamorro et al. [379] have developed guidelines and templates.

An important part of *designers' knowledge* [86, 129, 169] is *educating new designers* as a strategy. A subset of scholarship has explored interventions to increase designers' knowledge on ethical design to avoid the implementation of manipulative designs. Some concepts around this strategy are including ethics in the curriculum of UX programs [86, 166], promoting more ecological ways to engage with ethics, promote self-reflection, as well as promote ethical reasoning and intentions-oriented assessment of technologies when teaching design [167]. For this, Nelissen and Funk [324] use "ChoiceBox" a critical design artefact that would help designers to avoid manipulative designs in the privacy domain. We believe some of these concepts are also applicable to the strategy 'increasing designers sensitivity' as a way of increasing *designers capacity strategy*. Lastly, to complement this strategy of *designers capacities*, scholars have discussed the strategy of *promoting and integrating ethical frameworks*. Papers discuss the use of ethical frameworks like Value-Sensitive Design [180, 255], or creation of new ones that encompass manipulative designs [169, 306, 379]. Empowering designers within their organisations [379] is another suggested sub-strategy that we understand is intimately related to organisational practices and governance models.

In the third strategy, *designers responsibility* [169], scholars suggest the strategy of *design evaluation*, understood as giving tools and consequent accountability to designers within the implementation of manipulative designs. Thus, some concepts suggested are the use of design methods to test the presence of manipulative designs — e.g. A/B testing [314, 379, 45] — for self-auditing of designs [235]; as well as assessing the potential impact of the design [347], and looking at users vulnerabilities when designing [379, 240]. On a more concrete level Mildner et al. [306] suggest a framework for evaluation of ethical conversational user interfaces, Krauß et al. [246] suggest a framework for an ethical extended reality that can avoid manipulative designs, and Caragay et al. [70] suggest a framework that takes into account users expectations with a software development approach. Nimkoopai [328] also prototype a broad risk assessment tool for manipulative designs, while Shi et al. [403] provide and tests a risk assessment score with the same purpose.

HCI and Design Interventions for Users

This domain entails interventions in which technologists and designers alter interfaces (as the medium of interaction) to protect users. These interventions aim to address the interaction

Designers			
Intervention Strategy	Strategy	Concept	Prototype
Designers capacity — <i>Increasing designers possibility's to act upon designing without manipulative design</i>	Providing access to knowledge beyond data-driven design [1]	-	-
	Promoting ethical frameworks	Value-sensitive design Emotion assessment tool	Caragay et al. [70] -
	Increasing designers sensitivity	-	-
Designers knowledge — <i>Increasing designers' knowledge to prevent manipulative designs</i>	Educating (new) designers [458]	By engaging with ethical reasoning	ChoiceBox [324]
		Helping to focus on intentions when designing [167]	-
		With ecological ways for students to engage in ethics [167]	-
		Including ethics in the UX curriculum [86, 166]	-
		Promoting self-reflection [167]	-
	Promoting ethical frameworks	Value-sensitive design [255] Emotion assessment tool [1]	- -
	Supporting designers with knowledge	Automated tools	-
	Templates with ethical designs APIs [430] Warnings [45] Guidelines and best practices in ethical design	- - -	

Table 10.3: Summary of interventions belonging to Designers and developers domain.

				Frameworks for CUI [306] specific technology (i.e. conversational user interfaces [306, 246, 347], XR, [246], IoT [347])	
Designers re- sponsibility [169] — <i>Increasing designers account- ability</i>	Evaluating de- signs		Self-auditing		Kirkman et al. [235] and Chen et al. [80]
			A/B testing and UX methods for ethical design [314]		-
			Assessing and measurement of impacts		[403]
			Assessing users' vulnerabilities		-

Table 10.4: Summary of interventions belonging to the Designers and developers domain. (Cont.)

between the user and the manipulative design within the interface itself and constitute attempts to change the design of the user interface. There are three main strategies: triggering reflection, altering the choice architecture and other UI changes.

Triggering reflection. As Chang et al. [76] explain in their scoping review, a majority of studies use Kahneman's dual-theory [223] as the main theoretical lenses. The dual theory assumes the existence of 2 systems to process information — System 1 is automatic, and System 2 is reflexive. In HCI, it is applied to people's use of shortcuts and heuristics within their use of the interface. Hence, studies propose interventions to steer users away from the intentions of the manipulative instance by applying behaviour change knowledge to trigger the reflective system. Some of the strategies that have been proposed to trigger reflection are the use of goal setting techniques [285], friction [45, 316], seamful design [129] or cost reframing [409, 316].

Lukoff et al. [281] and Lyngs et al. [285] suggest goal-setting techniques in the interface to mitigate attentional harms. Prompting goals previously set by users aims to trigger reflection and prevent users from spending excessive time on video platforms [281]. Around the idea of reflection, some studies conducted on general populations discuss the use of friction and micro boundaries as potential countermeasures [316, 45, 78, 281, 255, 47, 281, 336]. "Educative nudges", reminders, and warnings are concepts also proposed as ways to increase reflection through friction [173, 162, 45, 270]. Long-term boosters [270] and more understanding of the

HCI/Design			
Intervention Strategy	Strategy	Concept	Prototype
Triggering reflection	Friction [45]	Educative nudges and warnings [45]	-
	Goal-setting [281, 285]	Reminders and goals	Lukoff et al. [281]
	Reframing costs [45]	Tool that provide costs along the websites*	Sin et al. [409]
	Postponement [409]		[409]
	Distraction [409]	Long-term boosters [45, 162]	[409] -
Choice architecture alteration	Changing language	Non-technical language [442] Caution with symbols [394]	[442]
	Homogeneous settings	Unambiguous multimodal cancellation protocol [390]	Device differences [453]
	Equal design options	Traffic light schema [34]	[442]
	Clear choices at the beginning of the process [31]	Better signalling of rejection [32, 390]	-
		Equal options (link and button) [31]	Berens et al. [31]

Table 10.5: Summary of interventions belonging to the HCID/Design domain.

Other UI	Imposing restrictions	Limiting articles in baskets* Removing infinite scroll as by default option* Blur screen with time*	
	Information [388, 389]	Providing more information	Schäfer et al. [388, 389], Kowalczyk et al. [243], Sergeeva et al. [396], and Habib et al. [183]
		Reducing information	Schäfer et al. [388, 389], Purohit et al. [366], Kollnig et al. [241], and Lyngs et al. [285]
		Features to validate information	
	Avoid visual cluttering Size Position		Schäfer et al. [388, 389]
		Equal options (link and button)	Berens et al. [31], Bermejo Fernandez et al. [34], Borberg et al. [46], and Utz et al. [442]
		Changing colours and defaults	-

Table 10.6: Summary of interventions belonging to the HCI/Design domain.

psychological aspect of interactions [237] also fall in this popular category of interventions.

Choice architecture alteration. Several studies try to understand the effects of the decisional space on the interface, or the choice architecture [432]. Thus, some of the strategies suggested are *changing the language* like reducing technical language from the interface [442], or being cautious and conscious about the use of symbols that might lead to confusion, specially in symbol-based languages [394]. *Homogenising settings* in all kinds of interfaces so users get used to them is another suggested strategy [390]. Lastly, *making design options equal* is a common strategy suggested by literature via different concepts: link vs button options should remain equal [31], rejection options need to be better signalled [32], using colour traffic likes schema to signal cookie options [34].

Other UI changes. There is a third strategy where prior work has tested interventions at the User Interface level (UI). Strategies in this space are using the appropriate size, providing information, [46], moderating visual cluttering [388, 389], and allowing autonomy in configurations and appropriation [396] to users. Within this strategy, Bielova et al. [38] experimented on cookie consent banners and included in their interventions a condition in which the risk of accepting was represented. Such risk representation proved to be effective in reducing cookie acceptance. Graßl et al. [162] also tested the effectiveness of “bright patterns” by highlighting options that are in favour of users’ interest in the context of cookie notices. Another strategy revolves around the information provided in the interface: providing information, reducing information, and including features to validate information are some of the concepts. Particularly, Schäfer et al. [388] test different visual countermeasures finding an interesting visual paradox to create countermeasures to combat manipulative designs. While users prefer to be informed, rather than exposed to more salient features, they also want to avoid visual cluttering when interacting with the interface.

Environmental interventions

This type of intervention acknowledges the multi-faceted aspect of manipulative designs and suggests solutions in which technologists and designers consider users’ environment as a mediator of interaction to protect users. These interventions look at other actors that mediate the relationship between users and manipulative designs: the social context and the technological affordances. While some strategies and concepts had been pointed out in the literature, we have elicited three strategies: leveraging context, leveraging social aspects, and learn from intertwined affordances.

When it comes to *learning from intertwined affordances*, Chaudhary et al. [78] advocate for interventions in the algorithmic system as strategy, suggesting concepts like refreshing recommendations to break the loop [78], or using varied recommendations as well as removing filters [463, 285, 433] and content from the newsfeed [366]. Both Lyngs et al. [285] and Purohit et al. [366] tested how altering the algorithm and filtering it would work in avoiding distractions [285] and time spent [366]. They observed that users feared missing information that is being removed [285] or content of their interest [366].

The space of *leveraging social aspects* was hinted by Westin et al. [458] when they sug-

gested empowering users against FOMO-centric design by accounting for the social aspect of privacy trade-offs and the need to belong [47] as strategies. When it comes to leveraging the context, Chordia et al. [88] consider the possibility of channelling users' fear productively in their interactions to avoid falling for the deception.

Organisational interventions

Some scholars call for broader interventions that target companies and their environment to avoid creating manipulative designs unintentionally due to their organisational structure. One strategy in this domain is *educating providers* [46], more concretely on the harms they can cause [46]. On a conceptual level, Aagaard et al. [1] explain how the "dark patterns badge" can be useful for companies to prove they do not contain manipulative designs. Another space relates to *structural changes and processes* (that also empowers designers within the organisation [150]). A sub-strategy in such space is to implement governance models for manipulative designs [379] that includes, for instance, better knowledge management within the companies to avoid designing manipulative designs [312]. These interventions are very related to co-regulation and self-regulation, since they are complementary. Lastly, a third strategy is *changing the incentives system in companies* [296] through combating monopolies [458].

Environmental interventions			
Intervention Strategy	Strategy	Concept	Prototype
Leveraging context — <i>Considering and using contexts to protect users within their interactions*</i>	Contextualising danger	Channeling fear productively [88]	-
	Physicalising risk* Understanding psychological aspects Cumulative experiences*	Reminders of past experiences of manipulation*, Reviewing cookies accepted in the past*	[38]
Leveraging social aspects* — <i>Considering social contexts and relationships of users to protect them within their interactions</i>	Leveraging social influence*	Using reviews-style to inform about MD [16, 215]	-
	Considering the need to belong [47]		
	Delegating*	Virtual agents for shopping*, Creating a service with limited clothes*	
	Controlling* Sharing*	Risky ball*, Showing others' algorithm *	
	Goal-setting	Reminders and goals	Lyngs et al. [285] and Lukoff et al. [281]

Table 10.7: Summary of interventions belonging to the Environmental domain. Interventions marked with "*" have been suggested during the ideation process described in the discussion.

Learning from intertwined affordances* — <i>Considering and using other technical affordances beyond the UI design to protect users within their interactions</i>	Changes in the algorithmic system	Refreshing recommendations to break the loop [78, 285] Varied recommendations [463] Lyngs et al. [285] and Purohit et al. [366] Filtering content [366, 285]
Increasing resilience* — <i>Supporting users in their process to recover from a harm</i>	External support (e.g. police)* Increasing self-efficacy* Normalising*	Tipline [230] - - Memes and public dissemination* -

Table 10.8: Summary of interventions belonging to the Environmental domain. Interventions marked with "*" have been suggested during the ideation process described in the discussion.

Organisational			
Intervention Strategy	Strategy	Concept	Prototype
Educating providers — <i>Increasing awareness of design impact caused by organisations</i>	Educating on harms [46]	Dark patterns badge [1]	-
Structural changes and processes (that also empowers designers [150]) — <i>Improving internal processes to avoid unintentional manipulative designs creation</i>	Governance models for manipulative designs [379]	Better knowledge management to avoid designing MD [312]	-
	Changing incentives structure [296, 458]	Shareholders agreement*	-

Table 10.9: Summary of interventions belonging to the Organisational Domain.

10.5 Study 2. Mapping the Experience With Manipulative Designs

10.5.1 Overview

Aggregating the findings from two qualitative studies and prior empirical work on manipulative designs, we outline users' experiences using a journey map. Originally a service design method, journey maps are used by practitioners to synthesize the findings of user research, convey key insights to the team, and find actionable solutions to identified pain points. The methods for data collection to use these maps are varied: note-taking [470], observations [294], critical incidents [61], diary studies [61], and physical prototypes [258]. In this study, we combine data from two different interview studies with users in situations of vulnerability and look at it from the perspective of temporal experience of manipulation.

10.5.2 Methodology

In this study, we developed an experience map of online manipulation experiences as an overview of the contextual and temporal aspects of the effects of manipulative designs. This experience map builds on the secondary analysis of datasets that we previously analyzed in published empirical studies. Secondary analysis of qualitative data is a recognised method that involves re-analyzing existing qualitative data collected for a different purpose [193, 210, 345]. Following O'Reilly and Kiyimba [345], we specifically talk here about re-use as our interview material is revisited and reanalysed by the same research team. Both interview datasets were aimed at understanding the experience of manipulation in a contextualised way.

The overall aim is thus not different per se, but the combination and synthesis under the form of an experience map for the present study involved looking at the data through a more temporal lens, and considering touchpoints and actors in a more analytical way. In the two studies — with young adults at social-exclusion risk (N=19) and teenagers at social-exclusion risk (N=6) — interviews combining different techniques, critical incidents, and probes with manipulative designs were used. In this study, we look at the data through a different lens, paying attention to the experience along the interaction and the relationships with other actors.

Data Collection

Both data collections were conducted in collaboration with an NGO working with people at risk of social exclusion. The interviews were conducted in the NGO premises, and we took specific measures to guarantee participants' rights and safety: information about the study, the consent form, and information about withdrawing was provided a month in advance. The interviewer also explained the consent form information orally before every interview.

The interview topics, divided into two sections, were: (i) experiences of felt manipulation and deception — including — what they do before and after such experiences and (ii) experiences with specific deceptive design patterns. A total of 19 young adults at risk of social

exclusion, and 6 teenagers at risk of social exclusion. Participants were fairly compensated, and the study obtained ethical approval from the University of Luxembourg. Participants' demographics can be found in detail in Chapters 7 and 9.

Journey Mapping

The purpose of this map is to create a visual representation of users' interactions and experiences with manipulative designs and underlying harms. Following the traditional structure and components of a journey map [225, 12], we report the findings through a temporal lens: from the time the interaction with the platform starts until users decide to cope with the consequences. Acknowledging the role of the context in the experience of manipulation, the journey map also includes the actions, thoughts, emotions and motivations, actors, environment, and harms that accompany every stage of the experience of manipulation. We report the findings using the same participant identifiers as the original publications for coherence purposes: pseudonyms for teenagers and anonymised participant codes for the young adults.

Analysis of the Stages of Felt manipulation

In the aforementioned studies, participants described different episodes in which they felt manipulated, deceived, and tricked by different elements of the interface design. We mapped these experiences along the following key phases: pre-interaction, interaction, and after-interaction.

By using open inductive coding, building on thematic analysis [51], the first author has refined the temporality of the interaction in more granular events and elicited associated elements — e.g., external actors that intervene within the interaction or the associated emotions. The rest of the authors took the codes from the first author and built similar maps to contrast with the first author. The map was refined through iterative discussions. To build blocks with the literature on manipulative design and have a comprehensive view of the experiences of manipulation, we incorporated some aspects already gathered in the literature to complement the synthesis of data with existing knowledge from the literature. While we are aware the authors did not think about the temporality of the experience when conducting their studies — with the exception of Gray et al. [163] —, we take their results and use them with the lenses of temporality.

User Stories

During the mapping process, we created user stories to exemplify ways in which individual, contextual, and temporal factors combine in the experience with manipulative designs. A user story is a plain English statement used in agile product development to articulate and convey user needs [4]. It usually captures the “who”, “what” and “why” of a product requirement. We use an adapted version of the standard Connextra format to formulate user stories: *As a (who wants to accomplish something), I want to (what they want to accomplish), So that (why they want to accomplish that thing)*. In the following, user stories have an illustrative purpose. They

are not exhaustive of all potential users need. They can serve as inspiration or a starting point for the ideation of solutions to address the problem of manipulative designs.

10.5.3 Findings

The experience with manipulative designs is presented in terms of stages of the interaction, users' actions, harms, and pain points (see Figure 10.5.3 for the visualisation). For clarity purposes, we structure this section according to the main stages that participants' experience with manipulative designs (encompassing the temporality prior to the encounter, during, and after the interaction). We identified six main stages in their experience of online manipulation: motivation for the interaction, initial assessment of trust, interaction with the manipulative practice, impact awareness, immediate correction and long-term coping measures. Every stage is further subdivided into actions. Noteworthy, not all stages occur in every interaction or for every user. For every stage, we include the interactions with the environment, thoughts and feelings that will lead to pain points that we consider drivers of vulnerability to manipulative designs.

Motivation for the Interaction

The participants' experience of manipulation starts with their motivation for using a certain system or platform. Participants do not aim to interact with manipulative designs, but they encounter them as part of their interactions with systems. The encounter with manipulative designs can happen in two ways: the user triggers the interaction with the system, or external causes act as the trigger for the interaction.

(1a) The user initiates the encounter with manipulative designs. Participants initiate the encounter based on either pragmatic or hedonic needs. Pragmatic needs — e.g., looking for a present on a website — are usually associated with a goal-driven interaction.

Hedonic needs are associated with psychological needs, and personal contexts; both play a role in how users initiate the interaction. Participants explained how emotions and feelings act as triggers for the interaction. *The need for relatedness*, expressed through feelings of loneliness and *lack of social support* channels the need to escape through social media platforms like TikTok or Instagram. Several stories by the young adults coincided with P9 testimony: "*I am lonely; I have nobody to share my life with; the mobile phone is my life*".

Young adults explained how they evade responsibilities by immersing themselves in social media or e-commerce platforms where they can spend hours following the infinite algorithmic-generated content. In this usage, participants encounter different types of manipulative designs — e.g., infinite scroll, choice architecture manipulation, or limited and scarcity cues — that will lead to some harm, without necessarily being visible to the user when initiating the interaction: attentional harm, or the reinforcement of emotional and psychological distress.

(1b) External triggers provoke the encounter with manipulative designs. In this case, the interaction with the system is initiated by an external trigger, which, for instance, takes the form of a pop-up window, a phone notification, a targeted ad, or an email. Contextual and

socio-technical factors mediate the interaction with the manipulative designs, e.g., economic incentives, emotions, and algorithmic targeting of common triggers.

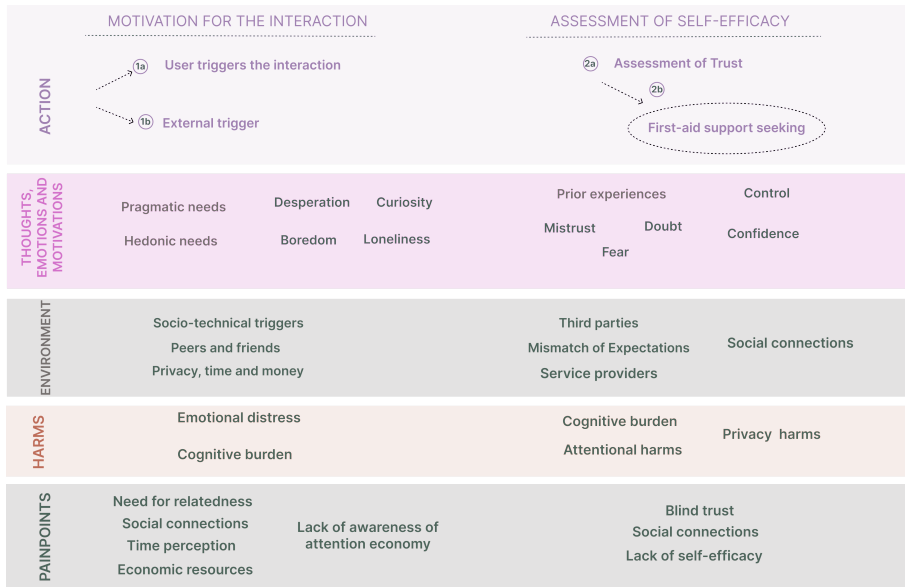


Figure 10.1: Experience map of the participants' experiences with manipulative designs (Part 1)

Economic incentives represent an important reason for participants to initiate an interaction with a system entailing manipulative designs. Participants were attracted by pop-ups, scam intents, or deceptive ads bombarding them with discounts and offers. These offer “free” services in exchange for their personal data and, ultimately, money. Teenagers expressed they would always prioritise free content or features, even at the expense of their personal data. Alex preferred to watch extra ads to avoid paying and would obtain pirated content, where he usually encountered manipulative designs. Some young adults explained how their experiences of manipulation started due to economic incentives, designs, and scams that tried to exploit that economic aspect. Gray et al. [163] also coincide in the socio-economic aspect of the interaction when their participants identified freemium products as manipulative, namely they would find an interaction manipulative when the interface asked them to pay, especially “when there is a mismatch between users’s expectation and reality: for example, instances where the freemium product was advertised as free to use” (p.11). This is the case for the participants attracted by free products that end up being deceptive.

Time perception and agency over time also play an important role in initiating the interaction with manipulative designs. Oskar would time the play-by-appointment video game to match it right after class. As explained in Sánchez Chamorro et al. [382], for teenagers, it was noticeable how their perception of time would shape their interaction with manipulative designs, and vice versa: they would feel they have nothing to do and therefore let themselves immersed in

systems using manipulative designs, but these sometimes ended up, in turn, shaping them. Similar experiences were shared by the young adults when explaining they would open TikTok during “downtime” (e.g., waiting for the bus). Yet, sometimes, their context was so shaped by the platform that they would forget they were not allowed to take the phone in class and would watch TikTok, as P12 explained.

Socio-technical triggers. Algorithmic profiling is another entry point for the manipulation experience. Although participants initiated the interaction with the platform, they explained how they felt manipulated, deceived, or harmed by content generated through profiling. For instance, targeted advertisements that promise free courses or jobs are directed to them because of the profiling by algorithms and platforms. Specific content that participants saw on social media platforms perceived as harmful — i.e. content from gangs — comes from profiling and targeting some participants as low-income users. Similarly, the content from influencers that participants felt was harmful may be due to the algorithmic targeting, although participants might not always be aware. P12 explained that although the content might hurt her, as it triggers her insecurities, it is not bad to see since “*those people also exist in real life, and you do not close your eyes to avoid looking at them*”. In our view, this conception of the content comes from a benevolent lens since the way social media platforms display content is not “democratic” but targeted. Thus, although participants may not be aware, some harms already arise at this stage, like attentional and psychological harm.

The nature of social networks inherently carries a *social component* as a trigger of the experience of manipulation, in which users incite each other to start the interaction. Participants indicated how some users trigger each other and foster the interaction with manipulation. However, this social aspect is not only tied to social media platforms. As explained in Sánchez Chamorro et al. [382], for teenagers, for instance, the role of peers and friends in triggering interactions was present in videogames or in yielding privacy. Also, Westin and Chiasson [458] pointed out the importance of the social aspect of privacy in these manipulative interactions. P13 explained how she starts a loop of time-spending on the phone because she sends things to their friends and vice-versa. Bösch et al. [47] also explain how the need to belong as a social need can be exploited in malicious interfaces on the privacy domain.

At this stage of the journey, we elicited that the *need for relatedness and social connections, economic resources, time perception, and lack of understanding of attention economy affordances* are pain points that act as drivers of vulnerability to manipulative designs.

User story 1: As I need an article of clothing, I go to the website of the online retailer [unaware of how the profile-based algorithm will lead me to spend hours shopping on their website]

With User Story 1, we see how participants are unaware of the profiling and algorithmic system that will make them spend hours on the platform with similar items. The time perception that participants have in such tasks may contribute to their endless continuation on the platform, which, added to the user’s economic resources, may increase the likelihood of staying.

User story 2: I encounter an ad for a job [which was targeted to me due to my joblessness] which convinces me to give my contact details [as I feel desperate to get a job] [unaware the privacy harms this will entail]

In User Story 2, we observe how the way participants are profiled and their personal conditions will make them a better target for specific manipulative practices that will make them yield their data. Privacy harm is, therefore, likely to happen in such circumstances.

Stage	Driver of Vulnerability/Pain point
Motivation for the Interaction	Need for relatedness and social connections Economic resources Time perception Lack of understanding of attention economy affordances Need for stimulation
Assessment of Self-Efficacy	Lack of self-efficacy Trustworthiness & Blind Trust Social Connections
Interaction and General conclusion	Feeling of control Lack of self-efficacy
Impact Awareness: Cease of the Interaction	-
Immediate Correction	Lack of self-efficacy Lack of resources (Mis)Understanding
Long-term Coping Measure	Lack of resources for resilience Powerlessness Normalisation

Table 10.10: Summary of Vulnerability Drivers

Assessment of Self-Efficacy and Trustworthiness

Before the interaction, participants assess their ability to execute the task overall, considering executing it without experiencing any unexpected consequences. Thus, when participants reported having suffered a scam or a deceptive interaction, they all felt capable of conducting the intended task, but they did not detect any signs of a scam.

In addition to this feeling of self-efficacy, the system's trustworthiness is appraised and impacts the rest of the journey. Thus, the perception of self-efficacy is strictly related to users' trust in the platform and the mismatch of expectations they perceive. Ultimately, if users do not feel capable of conducting the task, they resort to asking someone in their surroundings for help.

(2) Initial assessment of trust. As part of that self-efficacy, there is an important relationship between trust and control. Participants will keep the interaction going if they trust the system, which is a phenomenon already explored by prior studies: there is a relationship between trust and felt manipulation [163, 15, 453]. There are different reasons why participants trust the platform when encountering manipulative designs and, therefore, increase their self-efficacy: UI design and platform affordances, mismatch of expectations, referrals, and familiarity with the service.

Participants explained how they explore the platform affordances to identify potential triggers of mistrust that do not match their expectations. Mildner et al. [306] also explain how a key to (de)crease trust in participants is the mismatch of expectations. For instance, when P19 identify a disguised ad as fake, she does it building on her expectations.

"P19: It's fake news, you can notice by only looking at it, it is like..."

Interviewer: Why do you say so?

P19: To start, on the news...it's like they are not going to put a papaya, and a man like...I don't know. I would say that's fake, I wouldn't be bothered in looking at it because I feel it is a fake ad. And that's it, as I said, they are very common on Facebook"

Similarly, P15 explains she goes to the reviews to determine whether she trusts some e-commerce platforms. When P18 identified that her friend had been hacked, she realised that the posts from the account did not sound like her. P15 also look at the content and profiles of accounts that look suspicious to identify if they are bots or not. Those indicators increase the mistrust of participants. This relationship with trust is also explained by Gray et al. [163] as an important aspect of felt manipulation. Users identify signs on the platform that relate to the details of the platform design: advertisement or "sketchy" aspects raise mistrust on users [163] (p.9). Gray et al. [163] also pose a higher importance on trust in the first parts of the temporality of the manipulation — i.e. "initial judgement" and "inspection of details" — that serve as inspiration to develop this stage further. This idea is supported by other studies that explain how better perceived navigability on the websites relates to perceived honesty [447], and how the appearance of the interface increases trust [286]⁶

Participants' trust on the service also has a social component attached to it since seeing family and friends who suffered from scams or other types of emotional or economic harm would contribute to users' mistrust. Similarly, participants would also ask for referees to trust a service. This social aspect as a mediator of trust conditions users' future interactions, which was also pointed out by Babaei and Vassileva [16], and the Competition and Markets Authority [132] when they explain users increase their risk online because they are likely to trust strangers.

(2a) First-aid support seeking. Participants would seek help from their immediate surroundings if they did not feel capable enough to conduct the task they were aiming for. Thus, P9 explained that a friend had to teach her how to get flights on [the platform], because it was

⁶The study of M. Bhoot et al. [286] provides limited information about how they construct "physical appearance of the interface" as measure, but we still believe it is an initial indicator that, accompanied with other literature, may be useful to understand the experiences with manipulative designs.

complicated. While this is caused by their abilities and familiarity with the online domain, it is also fueled by the existence of manipulative designs and their previous experience with them. P4 explained how he finds it very difficult and confusing to get flights on Ryanair, having bad experiences with it, so he only goes to physical agencies to get flight tickets.

At this stage of the journey, we elicited that *lack of self-efficacy*, *blind trust in a service provider*, and *social support* are pain points that act as drivers of vulnerability to manipulative designs.

User story 3: [Since I have low feelings of self-efficacy] I am scared to continue ordering my flight tickets online [as I have experienced financial harm in the past] and will ask for help from my partner

In User Story 3, we see how the perception of *lack of self-efficacy*, added to negative past experiences with the platform, requiring participants to ask someone to support them. On the contrary, User Story 4 shows the opposite. As they never experienced a problem and *blindly trusted* the platform, participants would yield their data unconsciously.

User story 4: [Since I never saw any problems in the past with these types of ads], I trust the ad [unafraid of any negative consequences] and decide to click on it [unaware of the privacy harm that will entail]

Interaction and general conclusion: agency over the interaction with manipulative designs

The consequence of trusting and evaluating self-efficacy to the point of getting some help is that participants reach a general conclusion about the platform and a decision about whether the interaction may carry harm, reaching stage three (3) of the journey. In this part of the journey, users have a feeling of control over the interaction, expecting it to happen in a specific way. They expect to control the outcomes of the interaction, especially because they assume businesses have a specific agenda and they just have to deal with it. In this way, P12 equates protecting yourself online with going to the supermarket since “*if you are aware [of the manipulation tactics], it’s like in the supermarket, they put stuff on your face, so you go with your list, and you stick to it [to avoid buying other stuff]*” (P12).

When participants explain how they experience manipulative designs with altered choice architectures — e.g. pop-ups that nag you with very small crosses to close —, there is a general tendency to believe that, as long as users do not want to, those kinds of techniques have no power on participants. “Being strong-minded” (P9) and “not naive, but going on my own” (P17) give them the idea they have control over the interaction. This idea is supported by Bongard-Blanchy et al. [45] when showing how users express concerns about others falling for these designs without seeing the risk for themselves. We have seen this feeling very present in social media platforms, in which participants believe they are in control of their usage until

they realise they have spent much more time than they had foreseen — in the next stage of the journey.

The need to feel in control has been reported in some previous studies in which participants claim that if they do not have enough information during an interaction, they would not feel in control [366, 389] or they would believe the platform is deceptive [347]. Shamsudhin and Jotterand [399] theorised about how manipulative designs give a false sensation of agency that we see at the stage of the journey. We claim to be false agency because users believe they are in control of the consequences and harms, even when participants are not aware of those harms. For instance, when participants say they accept cookies because they trust the website providers, they feel in control, but the privacy harm is already done. Similarly, P1 and P5 explained they realised something was manipulative when they finally were asked for money through emails. It was only at that moment that participants realised it was not for free, but they had already given their e-mail addresses and personal data because they trusted to be in control of what they were doing. Thus, in privacy harms, this feeling of agency becomes especially relevant since habituation to the platform without seeing any tangible consequence may reinforce the idea of trust and, as a consequence, the idea of agency over the interaction, which can cause participants to be more prone to fall into these interactions. For instance, Alex explains how he would always accept cookie consent banners because otherwise, he cannot access the platform, and he does not perceive that it affects his privacy.

To gain this control, we see how some participants claimed to develop preventive strategies to avoid the effects of manipulative designs, like paying extra attention to what they read. Yet, they declare to find it hard to comprehend and often rely on the heuristic of trust in the provider and trust in themselves.

There is another tendency among participants to adopt resignation as a motto. As they believe there is nothing they can do, they will not try it. This was very present in privacy settings and cookie banners: all kinds of participants assert they cannot understand technical content, and they concede to the platform. Annoyance plays an important role here. Some manipulative designs annoy users by repetition, and wear them until they accept.

At this stage of the journey, we elicited that *feelings of control* and *powerlessness* are pain points that act as drivers of vulnerability to manipulative designs.

Impact awareness: cease of the interaction

At this stage of the journey, participants realise there is an impact that comes from the interaction, which does not necessarily imply they recognise the manipulative design as the source of harm, and cease the interaction. This realisation of harm is caused by the fact the interaction did not work result as intended — products did not arrive, money was taken not on purpose, a lot of time has been spent on social media, and a general feeling of discomfort after social media consumption. Thus, being aware of the harm will lead participants to make sense of what has happened, which will ultimately lead to immediate corrections.

Different studies support the idea that users are aware of manipulative designs when they perceive them to be harmful. M. Bhoot et al. [286] explained frustration is related to manipula-

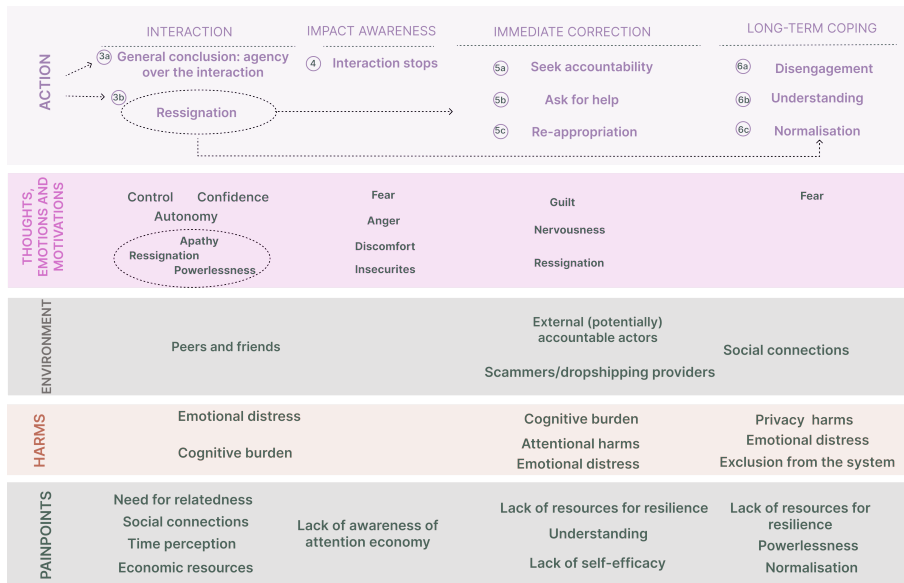


Figure 10.2: Experience map of the participants' experiences with manipulative designs (Part 2)

tive design identification. Gray et al. [163] identify, as part of the temporality of manipulative designs, “undesired interactions” and “negative results from interaction” as stages in which participants feel manipulated when they have “experienced negative consequences or being harmed by interacting with a manipulative user interface” (p.13).

Participants barely detected privacy harms with consent banners, nor with some pop-ups that they associated with the normality of the internet and marketing strategies, making them feel more in control of those. However, the harms of social media are very present: participants declare feeling addicted and finding themselves in a miserable position after exposure to these designs. When it comes to financial harms, they are perceived by users. The idea of the tangibility of harms already pointed out in Sánchez Chamorro et al. [382] becomes necessary because this will help to make sense of what happens and determine the best immediate correction strategy. When participants make sense of the harm, they might blame themselves as responsible in which the feeling of guilt prevails, but also feel others are responsible and, as a consequence, try to remediate. Feeling guilty will lead to taking some actions, while the feeling of resignation towards the harm will lead to assuming and normalising it as inevitable — in the next stage. In this stage, the harm becomes visible to participants; it is therefore possible that if users do not realise the harm, they directly pass to the stage of normalisation, which we also understand as a driver of vulnerability. User Story 5 shows that once there is a mismatch between what the participants expected and what they saw, the participants realise the harm and cease the interaction.

User story 5: Since I realised they were after my money, I [feared financial harm and] did not include my bank account after registering

Immediate correction

Once the harm has been noticed, and participants have made sense of it: giving it meaning and attributing someone accountable for it — themselves or others — there is an initial reaction of immediate correction of the harm. Participants who feel resignation towards the harm and assume that it is bound to happen — because of the system or because of themselves — will not react, starting the path of normalisation. On the other hand, some participants try to cope in some way — regardless they feel responsible for it or not. Depending on who they blame, how they perceive the interaction happened, and their own set of resources, they would opt for different ways of immediate correction.

Participants *seek accountability* to those actors they consider accountable and able to help them: platforms. Thus, participants go to the reporting systems from platforms. Similarly, the social aspect recovers importance again. Participants *ask for help* to friends, family, and peers, to recover from the harm. This was very common in teenagers who would use their parents or siblings to cope with the harm, ask them to hide their devices, or use them as a distraction from technology. They also seek help from other institutions like banks or other platforms to recover from harm. Lastly, participants *re-appropriate* technology leveraging their own resources with all kinds of creative solutions: using the platforms for venting instead of as a social application, trying to take a benefit and reverse the damage that had been done, or trying to “trick the tricker”. For instance, P1 explored how to take economic benefit from a platform after she was deceived into a subscription. *Deactivating notifications* and switching off devices have become a very common technique as an immediate correction measure.

User story 7: Since I am aware now they took more money than I expected and felt frustrated that I could do nothing about it, I decided to [reappropriate the situation and] “exploit” the system

At this stage of the journey, we elicited that *lack of self-efficacy*, *lack of resources*, and *(mis)understanding* are pain points that act as drivers of vulnerability to manipulative designs. In User Story 7, we observe the effect of resources to recover, as participants try to re-appropriate the situation when they realise the harm is irrecoverable. Conversely, in User Story 8, participants feel resourceless and seek help from those who might be accountable for their harm.

User story 8: Since I am aware now that I lost money because of the deceptive design and felt there was nothing I could do about it, I asked my bank for help

Long-term coping measure

After first reacting to the harm, participants adapt their behaviours in the long run according to these experiences. While some participants showed how they leverage their resources in the immediate correction, they feel more limited in the long-term

While participants leverage their resources for a first immediate correction, in the long run, participants feel more limited. Participants who felt that the ill-intended interaction was their responsibility would show the will to try to *understand the platform better* by paying more attention to future interactions and being more cautious. On the other hand, some participants would see their self-efficacy reduced after experiencing harm that they believe to be their responsibility. Hence, they would opt not to take services like that anymore, *disengaging* from the platform. Several participants explained how they would not reserve flights online or buy clothes after feeling manipulated by e-commerce websites with the prices or scammed by companies. Regardless the responsibility, *normalising* the existence of these mechanisms is another long-term coping measure, when participants believe there is nothing they can do to change the system, they would just assume the continued existence of these practices.

User story 9: Since I am aware that the last time I used the online platform to order something, I lost money, I disengage from online shopping [as I am afraid of losing money again] [unaware this excludes me from the system]

At this stage of the journey, we elicited that the *lack of resources for resilience* and the *feeling of powerlessness* are pain points that act as drivers of vulnerability to manipulative designs. Thus, in User Story 9, we observe how the lack of resilience resources makes participants disengage from the platform in the long term and, therefore, potentially suffer from exclusion from the digital system. On the other hand, with User Story 10, we show the effects of normalising these practices caused by overconfidence to control specific interactions, as they control other interactions offline.

User story 10: I don't think manipulative designs are bad [as I normalised them, comparing them with other practices such as marketing], and I feel confident that as long as I know what I want, nothing will affect me [unaware that they warp my autonomy]

10.6 Discussion

In this section, we first discuss what the temporal experiences of manipulation can tell us about the vulnerability of manipulative designs and what it means for design theory and practice. Later, by putting together the review of interventions and the elicited drivers of vulnerability, we also highlight design spaces and interventions that account for different drivers of vulnerability and elicit the associated challenges for designing counter-interventions. We

lastly discuss what mapping intervention spaces of manipulative designs means for future scholarship in the field.

10.6.1 Understanding Vulnerability to Manipulative Designs

By studying the temporal experiences of users in situations of vulnerability with manipulative designs, we have been able to elicit experiential pain points that contribute to such vulnerability. Malgieri [289] understands that certain circumstances of power imbalances can drive users to vulnerability, and such circumstances are multifaceted. Indeed, with the experience map, we have seen pain points related to socio-economic and socio-technical structures — i.e. economic resources, lack of understanding of attention economy affordances —, psychological aspects — i.e. the need for stimulation —, and experiential aspects — time perception, blind trust, the feeling of control, lack of self-efficacy, lack of resources for resilience —, and all of them seem relevant and intertwined to understand the multifaceted aspect of vulnerability to manipulative designs. Understanding the experiential aspect of vulnerability, provided the opportunity to unfold *experiential drivers of vulnerability* as concretely situated experiences that mediate the interaction between that user and technologies containing manipulative designs, putting the user at higher risk of the harms of those manipulative designs. These are not necessarily tied to a particular population but to an experience. Therefore, we hypothesise that such drivers can be translated into other populations that share the same experiences.

While some drivers of vulnerability cannot be addressed by interaction design, those that are related to the experience are directly related to how we design experiences, and therefore addressable by the design and HCI community, — both researchers and practitioners —, and contributing to what Barta [25] coined as technology-perpetrated vulnerability. These experiential drivers of vulnerability add a layer of nuance to the design complexity that designers face when aiming to create ethical technologies [164, 83, 84].

The added layer of this design complexity is visible in the relationship between trust and feeling of control — or lack of them —, within the experiences with manipulative designs. During the moments before the interaction, users conclude they trust the system and feel self-efficacy over the interaction and the manipulative design presented to them. Users feel pop-ups do not impact them even if they do not read. They also feel certain about the amount of time they will spend on social media and the amount of money they will spend. Users feel in control when they interact, giving them the “nudge” to interact with a design that can bring unintended consequences. Looking through the lenses of theories of motivation, such as the self-determination theory [20] — i.e. to motivate users, they need to feel competence, relatedness, and autonomy —, it is arguable that manipulative designs manage to give users that feeling of competence until they feel the harm. And this feeling of competence relates to trust in the service, or “relatedness”. Indeed, Kender and Frauenberger [232] explain that social media fulfils those psychological needs but with intentions that are not to promote the well-being of users.

The different studies that position trust as a motivator for interacting with manipulative designs [163, 286, 306, 232] give rise to a complicated paradox. Considering users might be

used to the platform and perceive no harm, users may trust platforms with manipulative designs [286, 163]. The aesthetics of these platforms also promote trust [232, 163, 286]. However, UX/UI designers will precisely attempt to induce trust in users to influence them [379]. Hence, services that manage to convey higher levels of trust will be more likely to convince users to fall into manipulative designs. Schäfer et al. [388] discusses this idea when testing some interventions to counteract manipulative designs. Indeed, in the results, we see that participants would accept cookie banners because they trusted Ryanair and Google. How do we break the cycle? Should we create mistrust so that users are not alerted about everything? How to design a balanced level of trust or mistrust to give confidence to users in the interaction while not alerting them excessively?

We see an intervention domain space that relates to this paradox. Thus, within the HCI/Design domain, some authors have argued in favour of friction for reflection [281, 316, 45] and other types of techniques of triggering users' reflective system [281, 285, 162]. While frequently discussed, there is still room for empirically testing such concepts to understand their implications. The works of Lukoff [281] and Lyngs [285] on agency can be useful for triggering System 1. Zac et al. [472], using an experimental setting, have argued about friction as a potential strategy to counter manipulative designs. With the problem of trust in mind, one can however not help but wonder whether this trigger of reflection is enough to protect users from manipulative design. Thus, users could be reflective enough and still trust service providers who use manipulative design.

10.6.2 Opening Solution Spaces Accounting for Vulnerability

Using the insights from the experience map, we brainstormed intervention ideas to elicit solutions against manipulative designs that account for vulnerability drivers. This exercise also allowed us to compare with the existing solutions and contrast the nuances that such a tool can provide. Using the drivers of vulnerability (illustrated by the user stories), we ideated a set of intervention spaces, strategies, and design concepts to protect users from manipulation by accounting for vulnerability. Looking at design interventions to manipulative designs through the lenses of drivers of vulnerability in a temporal aspect has allowed us to use the experience map for design reflexivity and open potential intervention spaces unexplored in previous works. Here we discuss some of the solution spaces and strategies that we unfold by looking at the contextual experiences of manipulation and that can target the drivers of vulnerability. We have highlighted a new environmental solution space with the following strategies: *leveraging context*, *leveraging social aspects*, *learning from intertwined affordances* and *designing for resilience*. We also want to mention the strategy *critical design* as part of the educational domain, and support goal-setting.

Leveraging social aspects

The relationship between a user and their social connections may trigger the interaction with manipulative designs. Although some studies theoretically point to it [458, 47], and Babaei

and Vassileva [16] look at the role of social proof in increasing users' interest in learning about manipulative designs, no published interventions are currently addressing the social aspect of manipulative designs. From a temporal point of view, social relationships as mediators of the interaction play an important part *before the interaction* (Stages 1 and 2) as well as in *coping* (Stages 5 and 6). Similarly, others could help to make users aware of the impact before it happens, contributing to cease the interaction (Stage 4). Therefore, leveraging how others mediate our interactions with manipulative designs can be an opportunity to rethink counter-interventions and design for protection beyond the one-to-one relationship between the user and the system.

How can we leverage the social aspect of manipulative designs to design counter-interventions? Some strategies we suggest can be delegating and controlling aspects of the interaction to others — i.e. involving someone else to decide within the interaction, or delegating the interaction to even an external party —; sharing throughout the interaction — i.e. showing your friends' algorithm to understand how it works on yourself — or leveraging social influence as Jafari and Vassileva [215] pointed out. They suggested that to increase interest in learning about manipulative designs, providers could mimic the review's look and feel, including information about the manipulative designs that products have. Designing for the social aspect can consider groups for designing in the context of manipulative interfaces. For instance, the knowledge gained in studies centred around designing for families [361, 81, 325] can help us rethink how we trigger the reflective system [223] by using social relationships as mediators. Thus, this social aspect can help to bring the experience of harm before the harm actually happens.

Leveraging context

This solution space arises from the understanding of the temporal aspect of manipulation, and a few of them have also been pointed out by some authors in their discussions: [232] suggests designing for well-being, and [47] understands psychological aspects of the interaction. We add the strategy of *contextualising danger by physicalising harm* and *providing cumulative perspectives when interacting*. Chordia et al. [88] already discussed the possibility of contextualising danger to avoid deception in safety apps, and we have merged with the idea of physicalising harm as an important part of the experience of manipulation since in some cases users are not aware of the manipulation until they see the harm. We also came up with different concepts that aim to change the experience of the harm by anticipating it: predicting how much money you are going to spend building on time you are on the website, providing the number of clothes that are still left to see to make them aware of the time they will spend, giving a counter about the time that has been spent on the platform, and explaining the harm on the cookie banner as Bielova et al. [38] suggest.

There are different alternatives to explore within these strategies. Combining the social aspect we anticipated earlier, a potential approach can be learning from the field of personal informatics applied to groups. Can families track their use of social media or their behavior when presented with cookie consent banners, and support each other? As an inspiration,

think about this fictional solution inspired by the field of personal informatics: a physical ambient artefact represents the time the family spends on social media. This representation is not individual to preserve privacy and avoid judgment, but collective to support awareness. The “risky ball” is connected to the social media consumption of one’s group of friends or family. The group decides on a time that they consider optimal to spend; when the threshold is passed, the ball starts to vibrate and progressively move. The higher the consumption, the faster and more erratic the ball moves. By physicalising the harm that others might be experiencing through the movement of the ball, users can help and support each other in stopping such consumption.

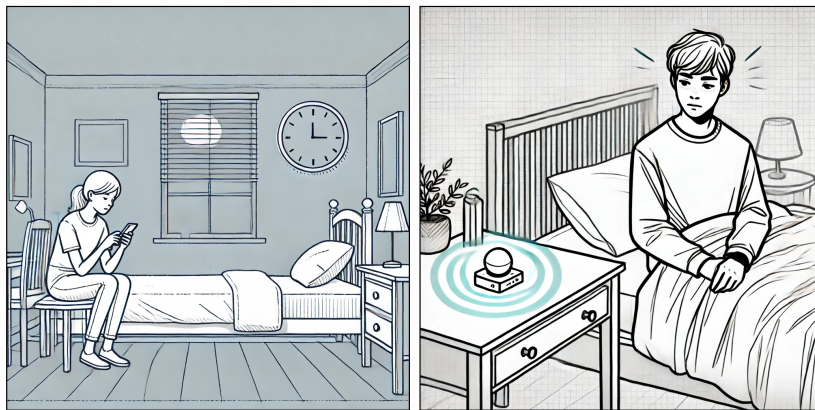


Figure 10.3: Risky ball concept. On the left, a girl is overextending time spent on her phone. The phone is connected to the “risky ball”, located at a friend’s home. The ball vibrates and moves, bothering her friend progressively as the movement intensifies. This image has been generated with the help of ChatGPT 4.

The presented illustrative concepts provoke a multitude of questions and broaden our horizons to other areas of interaction design. The risky ball aims to share risks through its physicalisation; it prompts us to question whether it is ethical to potentially “harm” other people in the process of mitigating manipulative designs through its shared component, but by extent also asks whether these people were not already involved in the harm from the start. Are manipulative designs a problem for individual users, that they have to resolve alone? It also prompts us to think about the potential perverse effects of our interventions: What if the artefact inspires competition rather than collaboration, or surveillance and negative feelings? What if a user aims to reach the “reprehensible” colour that comes with a higher use of the platform? These questions highlight different research directions that the community could investigate to support the creation of interventions.

We have also unfolded the strategy of designing for *cumulative perspectives* as part of understanding the importance of contextuality of the harm within the experience of manipulative designs. Thus, while other studies have looked at momentary experiences, by looking at the temporality of the manipulation and the relationship between users and harm, it allowed

us to conceptualise ideas to translate the experience of harm by thinking of the experience with a long-term perspective. As manipulation experiences would be cyclical and will impact the next interactions, trying to remind users about what they experienced on other occasions when they felt manipulated can help them prevent manipulation. Some concepts brought in this strategy are: starting the interaction with a survey that asks for how many clothes of the type the user is looking for they already have at home — which reassembles with the idea of reflection, understood as giving reasons for the interaction, that Sin and Munteanu [408] suggests.

Learning from intertwined affordances

Another important aspect to consider within the experience of manipulation is the relational aspect with the other technological affordances as actors within the interaction. In particular, socio-technical triggers play an important role in the motivation for the interaction (Stage 1) and assessment of trust (Stage 2).

Manipulative designs are not isolated but intertwined with other platform affordances; therefore, the experience caused by these designs is entangled. Some authors already suggested the need to target the algorithmic system as a strategy [230, 78]. Designing for the transparency of those socio-technical aspects, including, for instance, algorithmic transparency, also protects users against manipulative designs. Expanding knowledge and building bridges with scholars working on transparency and accountability of intelligent agents and algorithms can provide new insights into how algorithmic experiences overlap with manipulative design experiences [211]. Increasing transparency can help with the motivation for interactions and their realisation of self-efficacy since knowing how the algorithm works might give a more realistic understanding and awareness of their own interactions. Aligned with the scholars that advocate to increase more agency through the algorithmic system [281, 78, 396], we suggested different concepts that go beyond transparency: allowing users to customise the algorithm — giving options between infinite scroll or other forms of filtering, and providing initial options before the interaction —; using different algorithm architectures for e-commerce — i.e. a swapping type, in which you can only choose between 2 clothes you have to swipe —; and using a plugin that feeds the algorithm with noise, rather than actual data.

Supporting goal-setting

The personal situation of users, as well as their contexts — loneliness, boredom, and time perception— mediate the motivation for interacting with manipulative designs. In the early stages, in which users decide to interact with some platforms with manipulative designs and assess how capable they are of using these on their own terms, supporting users in setting goals is worthwhile. This solution space proposed by Lukoff [281] and Lyngs [285] suggest goal-setting techniques to reduce time consumption on social media. Here, goal-setting is seen as a way of giving agency to the users and triggering their reflective system [223] within their interactions with manipulative designs. While they focus on social media, goal-setting

interventions can be extrapolated to domains such as e-commerce or video games, in which we have seen participants spending much more than they expected. Designing solutions already in the *motivation for the interaction* stage could be effective in protecting users, and it is therefore worthy of further exploration.

Designing for resilience

The realisation of harm (Stage 4) and attempts to cope with it immediately (Stage 5) is a fundamental part of the experience of manipulation in which the lack of resources for resilience was presented as a driver of vulnerability. Nevertheless, currently, there are no design interventions in this space. We see it as an opportunity to empower users, especially in situation of vulnerability, since it would increase their resilience as a counterbalance to vulnerability.

In Stage 4, participants presented recovery strategies in which they leveraged their own resources. Hence, looking at how they use those resources or re-appropriate the platforms after harm can contribute to a better understanding of how we can design to prevent such harm. A first step towards this is to enhance existing recovery mechanisms and adapt to real users' needs and situations of vulnerability. Acknowledging that hindering users from recovery is also a manipulative design itself — like privacy zuckering [174] —, there is room for designing for resilience in the field of manipulative designs. Prior work on designing for resilience can be helpful for the community. More concretely focused on manipulative designs, related to this recovery, Schaffner et al. [390] propose different modalities for account deletion and homogeneous language for all of them. This idea could be transferred to other contexts, in which seeing similar mechanisms of reporting could increase recovery from harm caused by some manipulative designs. Thus, the multimodal aspect when it comes to suggesting solutions for manipulative designs is a recurring theme among scholars [347, 243, 306, 390]. Adapting the countermeasure to the modality of the platform can help users to re-appropriate the technology for their own benefit.

The use of a traditional method in design, like the experience map, has been proven useful to unveil some nuances in the understanding of users and their relationship with manipulative designs, but also to create counter-interventions. Thus, while some quantitative approaches have been proven useful, we advocate for using HCI and design methods that can better leverage experiences when it comes to the understanding of manipulative designs. Thus, some previous work already called for both extending experience design and UX methods to understand better the experiences of users with manipulative designs [297, 381], as well as the need to investigate new methods for the study of vulnerability [340].

Critical Design

Although several studies advocate for the use of design methods to support regulators, enforcement, and designers and companies when they evaluate their designs, we believe the use of *critical design* as design space can help within the domain of educating users. For instance, Nelissen and Funk [324] conceptualise a design speculation to help designers to avoid

manipulative designs, but these techniques can be used to help raise awareness. Thus, there is an extensive trajectory of critical design artefacts [22], design fictions [28] and speculative designs [189] as design tools that help to challenge the status quo.

10.6.3 Mapping Intervention Spaces as Opportunity for Future Research

We structured our review of interventions against manipulative design by using common vocabulary related to the exploration of design spaces. In doing so, we strived to bring clarity to the areas where prior work resorted mainly to umbrella terms at different levels of granularity or scope. Rare attempts have been made in the past to structure or map intervention spaces addressing manipulative designs. One notable example is the matrix proposed by Bongard-Blanchy et al. [45], with intervention measures (i.e. what are the interventions acting on) and intervention scope (i.e. the purpose of the intervention) as intersecting axes. Future research on manipulative design can leverage our structure and suggested vocabulary to ease further explorations of the design and solution spaces against manipulative designs. Similarly, looking at the interventions suggested by the community, it is present a lack of boundaries and limitations along the design space that can hinder how the community protects users. Thus, during the process of defining domains, we found several of them that lacked a clear "accountable" stakeholder. For instance, while it is clear who is expected to regulate, it is still open who should educate users.

Additionally, thinking about drivers of vulnerability as design challenges makes the problem of vulnerability more tangible. While "designing for vulnerability" can be seen as a very broad concept that is challenging to design for, "designing for promoting self-efficacy" can be more easily addressed by the design community.

10.6.4 Limitations and Future Work

This work presents some limitations. First, with the aim to comprehensively depict the interventions proposed by the HCI community (and not only the scarce interventions that are designed and tested), our scoping review encompasses a majority of interventions that are discussed by researchers yet not designed nor empirically tested. This explains why many of these suggestions are at the levels of strategies or ideas rather than elaborated design concepts and prototypes. Our review, therefore, does not — and does not claim to — provide insights into the effects of these interventions or the factors that make an intervention more or less successful. As the field and body of knowledge grows, such reviews will be relevant to conduct in future work. Additionally, considering our review as a part of the exploration of intervention spaces, we included all interventions suggested in the reviewed papers, notwithstanding the level of depth of their description. An intervention idea evoked in a single line of text without details or rationale (including generic claims such as "we need more regulations to address this issue") has the same weight in our review as an intervention carefully described, argued, and contextualised. This is the reason why we provide numbers or percentages simply

as an indication but purportedly remain at a more qualitative level by illustrating the different intervention domains with examples from prior work.

Regarding Study 2, our mapping is based on two qualitative studies conducted with limited samples and in limited contexts. With no claim of generalisation or exhaustivity, these localised results provided a source of inspiration for uncovering new intervention spaces based on potential drivers of vulnerability, which tend to align with prior work. They also made the case for the relevance of journey mapping to bring a temporal and contextual perspective when ideating interventions against manipulative designs. Noteworthy, while the benefits and drawbacks of secondary analysis or re-use of qualitative data are debated within the research community, from an ethical perspective "it is considered respectful to participants to make the best use of data that are collected" [345]. Finally, within the solution spaces we propose, the ideas we brainstormed and storyboard we propose are limited in their scope, not always taking constraints into account and serve as inspiration to illustrate opportunities within these intervention spaces. We encourage the community to pursue the efforts of mapping of online manipulation experiences, with other populations and in different contexts, and propose new intervention spaces or uncover requirements and constraints to operationalise existing spaces under the form of design concepts and prototypes to deploy.

10.7 Conclusion

This paper presents the results of two complementary studies, shedding new light on users' drivers of vulnerability in their experiences of online manipulation. Our first study reports on a rapid literature review of 139 papers to identify the type of counter-interventions currently suggested (or tested for some) by HCI and design scholars against manipulative designs. In a second study, we rely on the re-analysis of two datasets to investigate the experience of users in situations of vulnerability with manipulative designs. Using journey mapping as a research synthesis tool, we highlighted the relational, temporal, and situated aspects that may lead users to vulnerability in their interaction with manipulative designs. Our findings suggest new or underexplored solution spaces to create counter-measures against manipulative designs. Through this paper, we contribute to HCI research and practice on manipulative designs by (i) providing empirical insights into the contextual and experiential drivers of vulnerability to manipulative designs, (ii) reviewing and mapping existing interventions against manipulative designs, (iii) opening opportunities to further investigate new solutions spaces against manipulative designs, particularly suited to populations at increased risk of vulnerability.

10.8 Chapter Takeaways

By understanding the contextual and temporal aspect of vulnerability to the harms of manipulative, this chapter has provided a new lens for rethinking counter-interventions to manipulative designs. With this lens, new intervention spaces have been opened in the socio-technical domain, for instance, which allow for more holistic and contextual solutions to protect users.

This chapter, therefore, showcases the value of looking at manipulative design relationships with users in a more situated way by paying attention to their temporal aspect. Similarly, it has offered opportunities to use the mediated aspects of vulnerability to manipulative designs as inspiration for creating design interventions.

A Transdisciplinary Conversation: The Disentanglement

Buzzy: "Okay, so what we gonna do?"

Flaps: "I don't know, what you wanna do?"

Buzzy: "Look, Flaps, first I say, 'what we gonna do?' and then you say, 'what you wanna do?', they I say, 'what we gonna do?', you say 'what you wanna do?','what you gonna do', 'what you wanna' - let's do something!"

Flaps: "Okay. What you wanna do?"

Buzzy: "Oh, blimey, there you go again. The same once again!"

Ziggy: "I've got it! This time, I've really got it."

Buzzy: "So you got it. So what we gonna do?"

The Jungle Book, 1967.

11.1 The Entangled Problem of Manipulative Designs

Experiences with manipulative designs do not happen in a vacuum; they are situated and contextual, and so are their effects. From the initial taxonomies to categorise specific elements [95, 295] that build on Brignul's work on patterns [55], Gray et al. [169] moved on to UX strategies, as more complex instances than UI patterns. This work ultimately evolved into unveiling a socio-technical aspect of such strategies. Manipulative designs work at stratified levels [174]: macro-strategies that distort autonomy, meso-strategies that subvert users' expectations, and micro-UI elements that subvert expectations. This reveals that manipulative designs cannot be understood merely as a UI element; they are entangled with the complex socio-material environment of the user.

The evolution in the discourse on manipulative designs around 2018 [339] suggests the existence of a first wave of methodological approaches within the field. As Obi et al. [339] and the literature review presented in Chapter 3 show, researchers initially focused on understanding what manipulative designs were, how they were used, and where they could be found. After all, studying the effect of a phenomenon requires first scoping and defining it. Therefore, the first wave of papers set the ground with analyses of artefacts that contain manipulative designs [176, 180], and provided the first taxonomies [169, 95, 473, 295] that contribute to the understanding of what manipulative designs were. This wave was led by initial empirical explorations with users [111, 287]. Researchers explored the isolated relationship between users and interface in a behaviourist fashion: there is an interface that must have an effect — behavioural or cognitive — on the user when interacting with it. This model has given rise to several studies on cookie consent banners and e-commerce (e.g., [31, 32, 162, 45, 447]) and other less prominent domains (e.g., [306, 245, 347]).

The technological complexity accompanying manipulative designs reveals the entanglement of relationships that occur within their creation, deployment, regulation, and interaction. Hence, an empirical examination of the experiential aspect of vulnerability to manipulative designs can benefit many stakeholders involved in the manipulative design ecosystem. I here refer to entanglement by building on Frauenberger's [149] ideas. The way technologies are embedded in our lives blurs the line between users, interactions, and their effects to the extent that designing technologies becomes designing people and societies, and all social and technical actors involved have a shared responsibility. Therefore, it is important to understand which actors could play part in the reduction of vulnerability to manipulative designs and, by extent, the corresponding stakeholders that can enable this: e.g., users, HCI/design researchers, practitioners, and policymakers. Unveiling the roles of these sociological and material actors will help these stakeholders scope their space for action and determine where their boundaries and limitations are.

In this dissertation, I have presented how the personal and contextual situations in which the interaction takes place influence the experience of manipulation. Users' lived experiences condition their understanding of and relationship with manipulative designs. Considering vulnerability as a lens gives a more comprehensive overview of the problem of manipulative de-

signs, as well as ways to unify communities in search of solutions. In this chapter, I review the studies I conducted as a whole and discuss how looking through the lens of vulnerability to manipulative designs has provided not only a deeper understanding of users' vulnerability *per se* but also a more nuanced understanding of how the field can have a role in the fight against manipulative designs. By considering the different actors who may contribute to reducing vulnerability to manipulative designs, I discuss what changes in knowledge-making in the community mean to different stakeholders from a practical standpoint.

11.2 Disentangling the Problem of Vulnerability to Manipulative Designs through Users' Experiences

The extent to which some interactions make users vulnerable to the harms of manipulative designs has been a rising concern among policymakers. The fact some users respond differently to interfaces with manipulative designs depending on their personal characteristics — e.g. age, health, socio-economic conditions, technical knowledge — becomes crucial to policymakers since they associate a higher risk and harm to these personal circumstances [132, 341, 478, 284]. For that reason, understanding vulnerability is a concern reflected in the recent legislative texts — i.e., AI Act, DSA, GDPR —, as explained in Chapter 6. Some scholars have pointed out differences in interactions among subgroups of users — e.g., highly educated vs. less educated [45, 280]. With this work, I support this trend of scholarship and policymakers in understanding how, beyond the labels associated with these categories of users, different experiences make users vulnerable to manipulative designs.

So what are the relationships between manipulative designs, experiences, and drivers of vulnerability? In Chapter 6, we explained how the ecological model of Bronfenbrenner [57] fitted into an experiential approach to studying online manipulation. The environmental systems in which the interaction occurs contribute to the way users perceive, understand, and interact with manipulative designs (addressing RQ1). According to such a view, vulnerability to manipulative designs “occurs when one or more factors operate, giving rise to specific contextual and situated experiences that make users vulnerable” (see 6.6. The subsequent Chapters 7, 8 and 9 explore the experiences with manipulative designs of different populations in situations of vulnerability.

In Malgieri's terms, vulnerability drivers are facets belonging to different systems — macro, meso and micro. Indeed, the OECD reported some of these specific vulnerabilities [340] and we expanded them in Chapter 6 and 10. As those drivers might belong to different systems, Chapter 10 entails an initial distinction of vulnerability drivers along the axes of experiential and experienced, both of which are represented in the studies I present in this thesis. What we saw here were some drivers of vulnerability that are experiential: they are concrete situated experiences that mediates the interaction between a user and technologies containing manipulative designs, putting the user at higher risk of harm. E.g., a social trigger in a game, such as seeing all your friends have the latest in-game item. In other occasions, the experi-

ence itself was not a driver, but the user's description of their experience still helped unveil some of the drivers of vulnerability they face when interacting with manipulative designs, giving us experienced drivers. E.g., a user describing how being alone for a prolonged amount of time leads them to increase the usage of certain applications. HCI and design scholars can, for instance, influence experiential drivers through the design of different experiences, while policymakers might concern themselves with non-experiential ones — e.g., market structures and asymmetries or socio-economic factors — that can be elicited through users' experiences. For the objectives of this dissertation that value the users' experiences as malleable by design artefacts, I focused on the *experiential drivers* that may contribute to such vulnerability, as drivers that come from the experience (addressing RQ2).

Social drivers: Need for relatedness and social triggers

Users' social ecologies play an important role in mediating their relationship with manipulative designs. Lack of social support, need for relatedness, and social triggers were the social drivers of vulnerability I identified in the experiences participants reported.

Experiencing *lack of support* in their interactions can make users vulnerable when they face situations they perceive as out of their control. This experience was evident in teenagers who seek help, especially when they feel they cannot cope with the situation, and young adults, when developing proactive and reactive strategies to resist manipulative designs. Indeed, one of the resources young adults leverage to recover is people: their friends and family. Similarly, older adults reclaimed a need to be heard when facing manipulation online. There is a common underlying need for support that, if not met, could lead to increased vulnerability. In Chapter 10, social support was very present in the experience of manipulation.

On the other hand, *the need for relatedness* makes social connections mediators of users' relationship with manipulative designs. Social connections increase the exposure to some manipulative designs like scarcity features or algorithmic personalisation conducive to financial and attentional harm, accompanied by emotional distress. This driver was further reflected in Chapters 7, 9 and 10.

The social aspect as a trigger of vulnerability was prominent in Chapter 7. It showed how the social ecology of manipulative designs is very present in teenagers concerning how some designs work in specific contexts. For instance, scarcity cues in video games presented a social and identity related aspect that was not present in adults in e-commerce. We can see this contrast when looking at studies with adults, Tuncer et al. [439] found a negative relation between hedonic user experience and the presence of scarcity cues; while Tiemessen et al. [433] found that fear-of-missing out was not present in the presence of the scarcity cues.

Providers playing with the social aspect and sense of belonging can easily contribute to vulnerability [458, 47]. Westin and Chiasson [458] already theorised this phenomenon in relation to privacy concerns, in which they bring the concept of participatory reluctance from Cassidy as cited in [458], understood as users utilising a platform for reasons far from their interest in the platform — e.g., "I would prefer to use another social media platform, but my friends don't use any of the alternatives, so I have no choice." Similarly, this social ecology

seems present in how interactions with manipulative designs are triggered, as reflected in Chapter 10

This social aspect contributes to users' exposure to manipulative designs and carries additional challenges for protecting users from the effects of these designs. Conversely, Chapter 10 explains how those social aspects presented as a challenge in Chapter 7 could be leveraged as a solution space to rethink resistance strategies. Designing for groups, friends, and family to help each other, as well as fostering the resilience mechanisms after harm has occurred, is a potential way of using this social aspect in favour of the user rather than their detriment.

Interaction drivers: Impossible trade-offs and (mis)understanding

In some occasions, users experience the interface puts them in an imbalanced position, and they can do nothing about it. Therefore, I have gathered *impossible decisions* and *(mis)understanding* as experiential interaction drivers.

Experiencing that the trade-off within an interaction is too high is another driver of vulnerability to manipulative designs. When interfaces put users in front of “wicked decisions”, in which they cannot afford the trade-off — e.g. impossibility of investing money to avoid privacy intrusion, rejecting a discount in exchange for privacy, or accepting attentional harms instead of financial harm — users are more exposed to the effects of the manipulative designs. Chapter 9 discusses the normalisation of excessive burden, where participants found it normal to be asked one time and another about their preferences. However, they sometimes accept to avoid being bothered by these practices. In this case, providers ask users to trade their data for their experience on the platform. There is a similar impossible trade-off when providers ask users to trade their resources, as we saw in Chapter 7.

As part of those impossible trade-offs, the technological affordances behind the interface play an important role. The OECD report of vulnerability [340] explains how online vulnerability does not come isolated by design or personal conditions, but the intertwined algorithmic systems contribute to it in the line of Helberger et al. [194] and Barta et al. [25]. Thus, we have seen some of these effects in Chapter 7 and 9, in which the socio-technical aspect of platforms mediate some of their harms coming from manipulative designs — i.e. infinite scroll, or personalisation of content in a variety of platforms beyond social media.

There is an underlying socio-technical exploitation in these experiences that contributes to higher exposure to manipulative designs and that sometimes reinforces the cycle since, as explained in 10, the learnt experience with manipulative designs will condition the next ones; they are not isolated in time.

When it comes to *(mis)understanding*, users experience different reasons they do not properly understand the interface. In some occasions, I was even able to observe instances where participants did not realise about their misunderstandings; this has been very present in Chapter 9. In this sense, several scholars on manipulative designs have called for “a fix” regarding the provided information of the users and asymmetries of information [297, 308]. I have already mentioned how when evaluating counter-interventions, users demand more information, believing that would provide them more agency, as it was explained in Chapter 9. How-

ever, in my view, some of the studies that explore the idea of providing more information look at the interface from the point of view of the platform (i.e. “what does this interface need to be more understandable?”) rather than looking at the experience of (mis)understanding. Looking at those experiences provides a more nuanced idea of how trust, social aspects, and normalisation of platforms are tied to it and, therefore, might open new opportunities to think about other solution spaces beyond information asymmetries. How can we design counter-interventions that fight the normalisation of manipulative practices? Some of the strategies presented in Chapter 10 could work in that trend.

Drivers related to agency: Lack of self-efficacy, blind trust, and time perception

Some of the difficulties that lead to interactions with manipulative designs and consequent harm are related to how capable and confident users feel about obtaining the results they expect from a specific interaction. I therefore gathered a lack of self-efficacy, blind trust, and time perception as experiential drivers related to agency.

Experiencing lack of self-efficacy to be safe online was a common thread in the different studies. While this feeling stood out in older adults when presenting their magic machines, as explained in Chapter 8, the way users perceive their ability to cope (or not) with manipulation poses challenges to protect users, as explained in Chapter 9. Thus, low-levels of self-efficacy may lead to normalisation of manipulative designs and their effects: users may internalise they can do nothing to protect themselves. This brings another challenge associated with protecting users: how do we help them to increase this self-efficacy?

Indeed, when users are asked about their perceptions of manipulative designs and what they would need to be protected, some demand more agency over their interactions, as we saw in Chapter 9, which is in line with previous studies [388, 347]. However, what their behaviours show is that they still fall for these designs given they have normalised their existence [183, 336], and feel in control of their own actions inside of a system they cannot change.

The relationship between trust as a promotor of self-efficacy and a feeling of agency when interacting with manipulative designs also gives rise to some drivers of vulnerability. Some literature highlighted the importance of trust in the experiences of manipulation: users would interact with the services they trust [163, 306], and tend to mistrust when they lose control [388, 347, 366]. As explained in Chapter 10, *experiencing* blind trust makes users vulnerable to very common manipulative designs and harms: when they have not perceived harm coming from the service, their trust would increase, giving rise to situations of blind trust which sometimes lead to privacy, or financial harms, as well as to normalisation of manipulative designs. In light of these drivers, designing to promote self-efficacy and agency in users reveals itself as a challenge for design.

The *perception that users have over their own time* can act as a driver of vulnerability, as explained above. Teenagers, in particular, tend to have fewer obligations and more free time than adults. They might spend time on social media or video games on weekends, holidays, or evenings. However, in occasions where it is unwanted to spend time on cell phones, it can be difficult for teenagers to get themselves away from those. Teenagers from socio-economic

backgrounds with less social support can have more trouble in doing so, and it can be even more complicated if they have difficulties relating to attention — e.g., Attention Deficit Hyperactivity Disorder. Thus, although the teenager might want to limit the time spent on their phone, and believe they can do so, it becomes tedious when interfaces use manipulative designs to hide time or make users lose control of time. Here, manipulative designs are triggering the different layers of vulnerability, that accumulate [283], in this case: age, socio-economical background, and health.

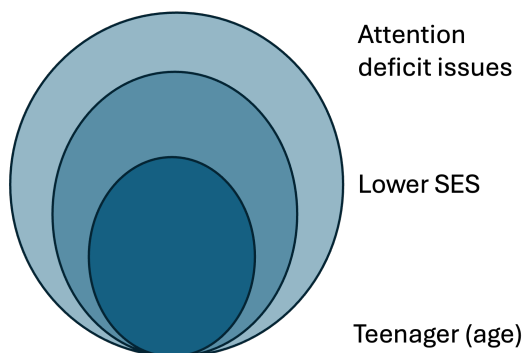


Figure 11.1: An example of different layers of vulnerability that are triggered in the presence of manipulative designs: a teenager from a lower socio-economic background who has attention deficit issues can rationally want to stop spending time on their cell phone, but might still not be able to when the manipulative design nudges them to continue anyways.

This thesis showed that users have an intuition of what is acceptable or not when it comes to digital nudging in their lives. Sometimes, participants seem to be afraid and reluctant after experiencing manipulation, which leads them to cope by disengaging from platforms. This means that other circumstances drive them to fall for the manipulative designs and prevent them from following that first intuition — e.g. feelings of loneliness. It is here that vulnerability occurs when other agents drive the perception and, consequently, the interaction that turns into manipulation. In the presence of manipulative designs, experiential aspects drive — in Malgieri’s lenses [288] — or trigger — in Luna’s approach [283] — vulnerability. Hence, an important conclusion of this dissertation is that vulnerability to manipulative designs can be present even when users (try to) act rational. Even when doing their best with their own resources, some users do not achieve the same outcome as other users in different circumstances. It is the experiential drivers nudging them to interact with manipulative designs that lead them to harmful consequences. This idea of the illusion of consent coming from the interface is related to this feeling of agency that users sometimes have in the presence of these designs and the normalisation of those, as it was explained above.

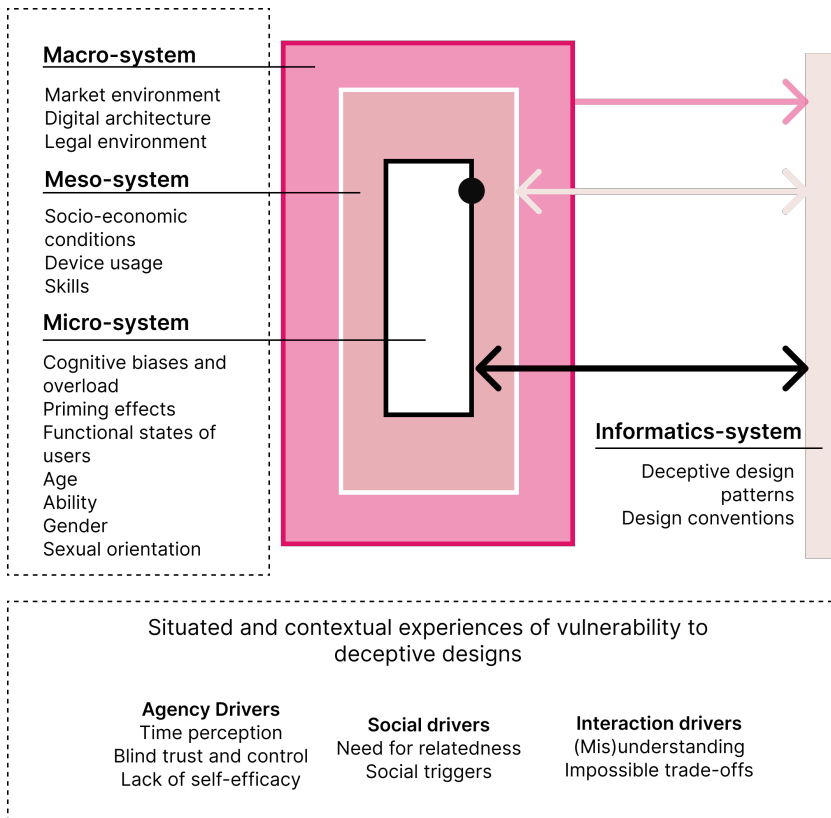


Figure 11.2:

11.3 The Disentanglement for HCI and Design Scholars

This dissertation diverges from previous literature in the way it approaches the problem of manipulative designs. Hence, the perspective adopted acknowledges not only that the interaction with manipulative designs is situated and that context matters; it also recognises that other factors mediate the interaction. The insights gained through this approach have unveiled new ways in which manipulative designs affect users: new harms and contexts arising in Chapter 7, different conceptions of manipulation in Chapter 8, and alternative ways of resistance to these designs in Chapter 9. Overall, this work contributes to broadening the understanding of what vulnerability means and to devise intervention spaces.

Chapter 3 discusses a need for alignment among the specific framing, objectives, and research methodologies used in manipulative designs research so that the community moves forward together. It is worth questioning previous paradigms and what it means for different stakeholders to propose a new paradigm in manipulative design research. Inspired by Harrison et al. [191], and in light of the results and methodological reflexivity, I here interrogate what kind of questions should be asked about manipulative designs and what are the under-

lying methodological implications. What are the elements that we, as a community, should move to the “center” of our efforts, and what are those that could be moved to the “margins”?

11.3.1 What Knowledge Should we Gather?

The necessary knowledge to better understand the effects of manipulative designs strictly relates to the purpose and meaning of HCI as a discipline. In Chapter 3, we called for alignment between disciplines to increase translational opportunities and methodological guidance to talk with other disciplines and make findings comparable. Similar multidisciplinary was discussed in Chapter 6, in which it was argued the use of experience design methods could help practitioners and legal scholars rethink how risk assessments can be conducted. The nature of HCI, as a discipline that combines computer science and psychology (among other disciplines), has fostered the search for epistemologies from psychology to complement the research in manipulative designs [201].

However, as aforementioned, vulnerability to manipulative designs is not always related to rational approaches. Hence, the entanglement of relationships between systems, affordances and users that gives rise to vulnerability in the presence of manipulative designs, calls to the need of expanding the knowledge to contextual situated interactions, rather than focusing on the cognitive-perceptual aspect — e.g. users perceive and interact with a pop-up. I claim for situated knowledge because the contexts mediate their interactions. Hence, interactions differ by types of service, moments of the day or personal situations; for instance, how often have you accepted a cookie consent banner because you were in a rush? Knowing about the temporal aspect of the interaction, the specific contexts and trade-offs during the interaction, the harms, and the different actors that affect or are affected by such interaction.

Looking at manipulative designs through the lens of vulnerability and its experiential aspects, this thesis understands manipulative designs as *elements of the interface that place users in positions of increased risk of harm*. Thinking about vulnerability to manipulative designs, therefore, comes with implications: the notion of *risk* indicates we should value the context in which the interaction takes place, while *increase* implies an underlying temporality before, during, and after the interaction to be taken into account. With this dissertation, I argue for shifting our center of attention from decision-making and choice architecture to vulnerability.

The *temporal aspect*. Understanding of the antecedents, correlates, and consequences of manipulative designs calls for a temporal perspective in our approaches, beyond the main paradigm focusing on momentary interactions. The *effects and harms* of manipulative designs are situated and contextual and call for in-situ approaches and the documentation of a wider range of contexts of online manipulation, beyond the widely studied cookie banners, e-commerce, and social media contexts. In our case studies, we gave participants the freedom to choose the context of manipulation they wished to comment on, either by leaving it entirely open (critical incidents and magic machines) or by offering a variety of choices (probes). This made sense when considering manipulative designs from a user- or harm-centered perspective. We want users to talk about their lived experiences, so inquiring about a context/platform they do not use is irrelevant. Sometimes, researchers will prefer to narrow down a study to

a specific context and population, which is not only an equally valuable endeavor but also necessary to have a contextualised understanding of the experiences.

This aspect also calls for a *systemic view to understand people's ecologies and contexts*. Beyond mentioning widely known field methods drawing from ethnographic or sociological approaches, our case studies brought reflections on the engagement of the researcher in the topic. For instance, the prior knowledge required to conduct and analyse the data of studies on manipulative designs. To prepare for our interviews, we got acquainted with the platforms that embedded manipulative designs. It felt essential to engage in inquiry and to know how to disentangle elements in the subsequent analysis. Lastly, as part of their ecologies, *the relationships* between different actors that are affected or affect the interaction are crucial to understanding the users' and rationale for the interaction. Thus, as highlighted in Chapter 7, manipulative designs also have network effects, and therefore, it becomes crucial to understand the set of relationships that intervene in the interaction with manipulative designs.

11.3.2 Which Methods Can Support Us?

In line with the search for new situated knowledge, it is important that the set of methods and tools we use contribute to meaningfully expanding such knowledge. While the combination of methods from computing science and approaches from legal scholarship is being used and proven useful for detecting manipulative designs and theorising about regulation, moving to situated experiences calls for different methods that can tell us about the more intimate relationship between users and manipulative designs. As stated in Chapter 3, there is a need for more methodological diversity and guidance to leverage the efforts of the manipulative designs community, which is one of the gaps when studying vulnerability in the digital domain that OECD [340] also highlights. Chapter 6 explained how the use of experience design methods could be leveraged to expand the knowledge about vulnerability to manipulative designs, as well as their countermeasures. On a more global level, I here present some reflections on the methods used in this dissertation to discuss which forms of epistemologies we can rely on for the study of manipulative designs, and provide some actionable points to expand the methodological toolbox within the community.

We conducted interviews using the *critical incidents technique* to understand felt manipulation in the presence of manipulative designs online in Chapters 7 and 9. Regarding manipulative designs, critical incidents are useful to understand felt manipulation as a proxy of the effects of manipulative designs, given their relationality [163]. Some advantages of this method to study manipulative designs have been presented as limitations in other methods, namely: (i) understanding the episodes of manipulation that have caused the most harm (from the users' viewpoint), (ii) studying potentially unfrequent events (depending on the internet use and type of use), (iii) gaining insights into contextual factors. Starting from the impact of the "critical incident", researchers can trace back elements of the interaction with manipulative designs. Considering how context-dependent these are, using a method that leverages contextuality is an asset. Our study with teenagers illustrates it: by discussing a situation that caused a participant emotional distress, we elicited a complex sneaking technique that added

levels of urgency in e-commerce and caused financial impact.

Some limitations of the critical incidents technique are common to interview methods in general. They rely on interviewee's memory and tend to highlight the most prominent or recent experiences. The interview guide is likely to prompt users on a topic. Factors of social desirability might also play a role, with either a desire to provide the interviewer with the information they seem to expect or to silence some experiences to avoid the embarrassment of reporting having been manipulated online. For instance, a teenager explained he experienced nagging techniques with pop-ups having a hidden close button (X) while watching movies online. He sometimes "failed" to close the pop-up and has to "endure" the forced advertisement because *he* failed. He said he does act to cope with it and should "*learn to pay more attention the next time*". Feelings of embarrassment, self-blame, and social desirability are frequently triggered by online manipulation. Reporting that one has been manipulated is sensitive, even more so when invited to report the most 'critical instances' and when the research focuses on non-normative collectives and populations in situations of vulnerability.

Interviews with Probes. The use of probes is a common method to gain insights into a particular phenomenon by showing artefacts to participants to react and reflect on. For manipulative designs, probes have been proven useful given the subtlety of the mechanisms that make users unaware of the influence [287, 45, 286, 111]. As a follow-up of critical incidents, we used interface screenshots including manipulative designs as probes to gain insights into the relationship between participants and manipulative designs. We selected the interfaces by contexts according to the ontology of manipulative designs by Gray et al. [174] and to represent a variety of mainstream platforms and manipulative elements.

The use of probes allows participants to reflect on the effects of those interfaces, by commenting on the specific context they find them in and in which specific ways they interact with these elements, sometimes in other contexts. This has been useful to document new contexts of online manipulation affecting specific groups. For instance, we showed teenagers e-commerce interfaces with nagging — continuously asking the user to make decisions they already took [174]— and changing the hierarchy of choice architecture [174]— but they recognised the strategy as very common in other contexts closer to their routines — e.g., offline fast-food kiosks and pirated-content websites. Without showing the probes and relying on critical incidents, we might not have elicited these contexts that add new knowledge to the community. Probes are also helpful with populations in which psychological harm may arise easily (e.g., low self-esteem). We observed how low-educated young adults experienced nervousness when asked about online manipulation using the critical incident technique. The subtlety of these mechanisms made participants feel embarrassed: a researcher was asking them about things they had not considered until that moment. Using probes of manipulative designs changed that feeling and created relief, as they recognised almost all the interfaces presented.

I must note that only relying on probes of manipulative designs has some important limitations. Directly showing manipulative designs to participants can have a priming effect, possibly leading to participants not reporting their actual encounters or perceptions about these

patterns. Thus, the researcher's ability to avoid priming into the aspects of the interface that can be manipulative is key if the intention is to elicit perceptions of manipulative designs. We thus suggested combining it with another method to triangulate data.

We also used *magic machines workshops* [11] in Chapter 8 as speculative co-creation methods, which place the participants in a "magical context" so they do not face the hurdles of talking about technology, while expressing their needs freely. In the context of manipulative designs, the low-tech and seemingly not-so-serious aspects of the method helped avoid the hurdles of interacting with a platform and, more broadly, talking about technology, which could be problematic for populations with less self-efficacy levels. As the method aims to empower users, participants are more at ease and less prone to social desirability biases. We aimed to shift the power dynamics and bring the magical atmosphere as early as the warm-up, asking them to choose a superpower. With older adults, we observed how the session's dynamic helped them talk about their experiences. Participants created a magic machine that protects them on their own terms. Through this group activity, participants recognised themselves in other stories and provide their perspectives.

This technique, applied as a user research method, tend to reveal major issues encountered by participants when interacting with manipulative designs, rather than specific issues at the interaction level. On some occasions, participants talk about usability problems rather than the effects of manipulation. For this reason, the onboarding of participants and prompt into the problem is key. While introducing manipulation as part of the workshop instructions might look like a limitation, we aimed to situate participants into moments of felt manipulation (and avoid having them build machines to overcome off-topic usability-related issues).

As explained in Chapter 4, these methods were selected as a way of trying to situate users into their own experiences. Based on the challenges to the study of manipulative designs effects on users, we argue that this second wave could flourish by striving for:

Bringing methods that count on lived experiences into the study of manipulative designs The HCI domain, as a multidisciplinary one, has incorporated a wide of methods that value lived experiences, especially in what is considered the 3rd wave of HCI and that comes from more traditional ethnographic inspirations [191, 123]. I therefore suggest the following practical considerations:

- Incorporating ethnographic methods into the methodological toolset of manipulative designs will help to contextualise the interactions that so far have been tested in a more behaviourist approach.
- Incorporating participatory design approaches can increase the value of those lived experiences since it will help to include the voices of those who are affected in the process of designing protective measures for themselves.
- Similarly, we can also incorporate communities to be integrated in the design of countermeasures by creating measures with a more user-centric approach.

- Researchers should be familiar with the platforms and interactions studied so they can properly understand the situated experiences.

Long-term evaluation paradigms, to understand the antecedents, correlates, and consequences incurred by manipulative designs.

Longitudinal studies that take context into account, we can learn from psychology, like ambulatory assessment, as well as trying to engage with real settings. In this sense, Moser et al. [316] explain how researchers “as defenders” do not have the same resources and conditions than companies “to attack”. Hence, taking a long-term perspective can contribute to have a more deeper understanding of the effects, harms, and contexts in which manipulative designs affect users in order to protect them better. Concretely, some practical considerations to take into account are the following:

- Consider evaluation of interfaces over time, even when conducting expert evaluations of design artefacts. Some examples can be found in Schaffner et al. [390] and Sheil et al. [401] or Gray et al. [170].
- Aim for understanding the causes and consequences of the users’ interactions, including their relationship with other actors that take part or can be affected by the interaction.

Conscious methodological choices around the “unawareness” of online manipulation, which is often pervasive and subtle. Online manipulation is a phenomenon that (some) users are likely to be unaware of. This potential unawareness is a key challenge, which calls for a reflective and creative approach to research. Related fields face similar challenges, in particular usable security and privacy [115, 116]. On the one hand, we can choose to prompt the participants about the specific context of manipulation. This approach has been dominant in the qualitative inquiries proposed by our community so far, and we do not deny its merits in investigating specific research questions. In our case studies, we initially tried hard to resist prompting participants to not influence the types of experiences they would share. In our instructions, we used the harms as an entry point and proxy for detecting and framing online manipulation, e.g., *When is the last time you did something you did not want to or did not initially intend to do on the internet?* While successful to some extent, our pilot tests have shown that the ambiguity of this prompt created confusion or led to off-topic conversations (e.g., on usability-related issues), which would unethically waste the time of the participants and the researchers alike. Eventually, we added examples to support framing the topic of the interviews or workshop sessions. In the probe study, we did not label the screenshots of interfaces as manipulative, and used the laddering technique to unveil the awareness of manipulation. Some practical considerations we elicited from the study, are the following:

- It is important to consider “manipulative designs” and “online manipulation” as research topics involving hard-to-reach populations [398], no matter the participants in the study. Talking about suffering from manipulation is a complicated topic and sometimes stigmatising, therefore, different methods of strategies should be used to minimise harm

to participants. At times, prompting the participants on the topic of manipulative designs triggered negative feelings of embarrassment, shame, or guilt in the aforementioned studies. Imagine how uncomfortable it can be to realise during an interview that you have been manipulated regularly without being aware? How can we acknowledge the sensitivity of this topic and "prompt" with care? Laddering the conversation from general topics to more specific ones can help the researcher steer the conversation in a more organic way. Similarly, being flexible and prepared for users' unwillingness to discuss specific topics, as well as being preventively alert to notice when participants might want to change conversations or activities, will help to create a more comfortable atmosphere to investigate the topic in-depth but under participants' limits.

- Attempts to study manipulative designs without bringing them to the awareness of the participants have been mostly made in experimental settings [38, 31, 439], and often reflected various uses of deception as a research mechanism, which also embodies ethical implications. For experimental approaches, techniques for hard-to-reach populations can be used, like the use of vignettes to elicit behaviours and attitudes that participants would usually not easily report [26, 60].
- In other domains, the "invisible" or "intimate" aspects that are hardly accessible to the researcher can be researched through first-person methodologies like auto-ethnographies [326, 466]. While they can be used to address some research questions to manipulative designs, these approaches do not support overcoming the "unawareness" challenge nor is the researcher usually representative of populations likely to be more vulnerable to online manipulation.

11.3.3 What Values Should We Stand Up For?

As a community, questioning the values underlying research on manipulative designs is fundamental. It helps understanding how the community should move forward, especially as we overall unite around the need to eradicate manipulative designs and work towards that direction [168]. As already mentioned as a core argument in this work, vulnerability to manipulative designs is mediated by the experiences and contexts, and even in some occasions rational: some contexts give users an illusion of agency, and they suffer the effect of manipulative designs even if they put all their capacities into avoiding it. Thus, the experiential vulnerability drivers give rise to contexts in which users are going to be necessarily vulnerable. Here, it is important to consider design research and practice a way of designing the world we want to live in [149]. If contexts mediate effects, design artefacts have an important role in those effects and, consequently, in users' lives. It is not possible to separate interaction designs from their effects, and therefore, consider that, in the realm of manipulative designs, design research and practice need to be a tool to protect users from such effects.

Justice as a value in interaction design. This dissertation introduced different design spaces. It is also worth mentioning the idea of "fair patterns" or "bright patterns" as a direct opposition to manipulative design. These patterns have been discussed by the community as design

patterns that would purposely nudge users into their best interests, trying to avoid harm [162, 384]. One can, however, still question the underlying idea: bright patterns are still manipulative “for the good”, which would remove agency from users. In the case users do not have the capacity to counteract the exposure or effects of manipulative designs, and design is a tool to design the world we want to live in, is it not the duty of design to provide such agency?

While it is not the purpose of this dissertation to offer a treaty of justice in interaction design — that has been suggested by several scholars [98] —, I believe it is necessary to reconsider justice as a value when moving towards a new paradigm of manipulative design research, especially in light of what vulnerability means for users. Thus, the situation of what design research should do to help users in situation of vulnerability could be interpreted through the lenses of theories of justice of Amartya Sen [395], which in design theory has been coined as the Capability Sensitive Design (CSD) approach [414, 214]. CSD is an approach in the interaction design process that allows recognising structural problems of persuasion on vulnerable groups. Defining the different needs that users expect to cover when interacting with the interface, will help designers inform their decisions during the design process from a proactive perspective.

The CSD, inspired, takes Sen’s Capabilities Approach, which considers justice a matter of choice. It is about the possibility of choosing what a person values in terms of doings or beings [214, 395] (“*functionings*”) and making them happen. Only when every user is allowed to choose their own set of “*functionings*” and translate them into “*capabilities*”, is it possible to talk about equal treatment. To translate “*functionings*” into “*capabilities*”, design needs to provide the adequate conditions. As an example, the functioning “*having a sexual orientation*” will not be transformed into the capability of “*expressing sexual orientation*” if a web form does not recognise all types of sexual orientation [214]. If users in situations of vulnerability driven by different experiences cannot properly exercise their agency over manipulative designs, meaning they do not have those same capabilities, design research and practice have a duty to restore the situation and promote justice within the interactions. A potential to do so is, therefore, fair patterns or bright patterns as a way of giving the capability to users in situations of vulnerability to have agency in the presence of manipulative designs.

Care as a value in research on manipulative designs. Due to their nature, specific populations might be more impacted by manipulative designs and likely more vulnerable and harder to reach when researching technology. Our case studies all engaged with populations at risk of vulnerability, which led to invaluable insights into the experience of manipulation that prior work did not address. We argue for a focus on care to conduct research that goes beyond observing generalisable causes. Indeed, using the layered vulnerability approach to manipulative designs proved helpful in understanding intersections of vulnerability along the matrix of domination. For example, in the study provided in Chapters 7 and 8, I uncovered drivers of vulnerability that intersected with age, but also socio-economic factors.

Some methods can help create a nuanced knowledge of the impacts rather than solely proving a change in behaviour caused by manipulative designs. To generate this knowledge, the study of manipulative designs not only requires knowledge of the contexts but also inter-

est and sensitivity towards the populations that are more vulnerable: they are more likely to be impacted and have fewer resources to recover from those impacts. The choice of methods should account for the more vulnerable populations because they are more impacted. As an illustration, while deception studies are worthwhile in many domains related to usable security (e.g., phishing interventions [114]), deception also entails the risk of harming research participants - a consideration that cannot be avoided when working with vulnerable populations and "practicing what we preach". Adopting the posture of a sensitive researcher [272], making efforts to include the voices of hard-to-reach populations, and seeking methodologies that balance the power between researchers and participants (e.g., magic machines workshops) should be key elements informing our methodological approaches.

Naturally, there is no one-size-fits-all approach to overcome all the challenges we shed light on. Reflecting on these considerations and making them transparent is an important step forward. While limited, the reflections we shared illustrate how to think of research trade-offs as well as our positionality as researchers in this field. By caring and tackling the uneasy, by making efforts to include hard-to-reach populations and acknowledging the vulnerability aspects involved, we can create a more robust body of knowledge on the effects of manipulative designs on users.

11.4 The Disentanglement for Practitioners

The extent to which users are vulnerable to manipulative designs is especially relevant for design practitioners as actors that share ownership and agency in creating manipulative designs. Chapter 5 unveiled specific problems that designers encounter when they want to steer users without the presence of manipulative designs. Designers show tensions in understanding when applying concepts such as manipulation and user agency, which they implement according to their own views and set of values. The tools they use and the standardisation of UX practices and methods (e.g., A/B testing to reveal vulnerabilities in users' decisions, or aggregated analytics) have led practitioners to normalise the exploitation of users' vulnerabilities. This ultimately contributes to normalising these practices among users, as explained in Chapter 9. There is a reinforcing cycle in which both users and designers see some design practices as permissible, agnostic, and neutral, as if they do not impact users. For instance, the presence of nagging pop-ups that increase cognitive burden and may lead to financial and privacy harm is widely accepted as a common marketing practice, as explained in Chapter 9.

Why should we break this cycle? Continuing to exploit users' vulnerabilities generates situations of injustice for some collectives. Following the conceptions of entanglement in HCI from Frauenberger [149], designers do not design for interactions; they ultimately design people and their effects. Hence, it is necessary to design for the world one wants to live in. This does not mean practitioners are alone or should have knowledge in disciplines such as ethics or law, but there is a need to support designers in acquiring such competencies [164, 84]. Thus, as explained in Chapter 10, taking into account the experiential drivers of vulnerability adds a new layer of nuances to the ethical design complexity that designers face when aiming for

ethical user interfaces as mediators of ethics with responsibility over the design outcomes [169, 164, 87, 86]: dealing with the paradoxes of trust, information and control puts designers in a delicate position.

There is a vast amount of literature to understand what practitioners need to make more ethical designs and consider beyond-human consequences [155]. Prior work also involved communicating better with practitioners and developing methods that adapt to the reality of design practice [85, 165, 175]. However, it is important to reflect and question the way UX is practiced at its very core, and to extend such reflections beyond academic conversations to include practitioners. Some issues with manipulative designs can be seen as the consequence of some past UX design trends such as persuasive design or behaviour change done without ethical reflection. If these practices are since being questioned by scholars, some practitioners may still think about persuasive design (e.g., from Fogg [142, 141], Cialdini's principles of persuasion [89], or nudge theories from behavioural economics [432]) as neutral or even harmless. Not questioning our practices as a community can make us partially accomplices of surveillance capitalism, as pointed out earlier.

Ideas have been proposed in the literature to cope with these matters: increasing designers' knowledge, capacity, and responsibility [83, 164, 169, 255, 167] are core strategies intended to promote manipulative design-free practices. Among these are reflections on how UX design curriculum [83, 164] and how we educate reflective practitioners equipped with a toolbox of ethical methods [167], as well as co-regulatory approaches that could support practitioners within companies. More concretely, Chapter 5 explains how the context of practitioners also matters in creating manipulative designs and points at some co-regulatory approaches that could help empower designers. By creating governance models inside companies that give voice to designers, companies would unite design, legal and ethical teams within the organisations, which would contribute to more comprehensive knowledge and to better document the design implications of specific choices. This may allow companies to empower ethical designers against the interest of higher clicks per minute, and to be more transparent with regulators. With this aim of a more transparent and conscious way of designing, I explained in Chapter 6 how some tools can support risk assessments for vulnerability in the development of technologies. Regarding more design-led approaches, designers can use drivers of vulnerability to protect users by leveraging their strengths but also using vulnerability as *material for design* [402] (p.7). In the same way that other works use embodiment [440], live streams [402], dreams [64], or sensors [402] as material to inform designs, vulnerability can be used as material by leveraging the intimate aspect of the users' experiences of vulnerability to protect them and avoid the stigma.

I want to echo here Helberger et al. [194]'s theorisation of digital vulnerability, which specifies that *"so far, under the (neo)-liberal market model, large commercial tech platforms have enjoyed ample room to shape the digital marketplace [...]. In short, addressing vulnerability and bringing fairness into the digital marketplace is not simply a question of empowering consumers, but of changing digital markets."* (p 196). Hence, the entanglement goes beyond the user and design responsibility notwithstanding; acknowledging it does not imply removing responsibility

from designers. For this reason, the discussion of vulnerability to manipulative designs cannot overlook this approach and suggest new horizons in such a line.

Yet, it is understandable and reasonable that scholars in manipulative designs often work under the belief they cannot change the system and that companies will likely not contribute to this change [347]. Again, Moser et al. [316] depicted this problem about the community, which lacks the resources to protect users in comparison to companies. However, I would like to invite us to reflect as a community: How to overcome adopting a defeatist position in this area? Unlike other communities, there is at least an agreement on how harmful these manipulative practices are.

11.5 The Disentanglement for Legal Scholars and Policymakers

As stated in the introduction of this work, policymakers are trying to protect users by regulating manipulative design practices. Particularly, Chapter 6 exposed the problems of regulating vulnerability to manipulative designs. Along this dissertation, we saw that the experiences of users reveal harms when interacting with technology, and that these can contribute to informing future designs that account for vulnerability. If the design community can account for users' experience, why would legal scholars not also account for them?

With the evolution of the digital space, regulators and legal scholars also experienced a change in their knowledge and epistemologies. In a discussion about how the internet could be or should be regulated or not, Lawrence Lessig institutionalised the idea that "code is law" [269]. Lessig explains that the technical architectures that make a product are already shaping users and the world, and therefore "regulating" the way we interact with it: code is also "law". Hence, regulators should not overlook that. In the HCI community, this idea has also been explored through behaviour change technologies. Lockton [275], for instance, explains that it is the digital architecture that affords users in specific ways, the one that influences behaviours and the whole ecosystem in general. This raises the question, if "code is law", and technological artefacts shape how the world works, why are users not "law" as well? Why are users not shaping standards, regulations, and designs?

Taking an experiential approach that accounts for the experience of users when facing manipulative designs is crucial for different reasons. For instance, when regulators take relational approaches in other words, the harms exist in relation to a user. Thus, the Digital Services Act (from now on "DSA") considers manipulative designs to exist when users' autonomy is distorted; hence, only users can teach us about that. The user point of view is vital to regulate manipulative designs if manipulation is about 'exploiting users' vulnerabilities'. There is a need to learn from an empirical point of view of vulnerability to tackle manipulation. Similarly, the systemic risk approach of the DSA, which aims to evaluate mental-health risks coming from platforms, which might be related to manipulative designs, requires a deep understanding of users' experiences with platforms.

Thus, as explained in Chapter 6, legal scholars and policymakers can learn from HCI to understand experiences. Thus, it is not only about taking empirical evidence from studies but also learning about the value of experiences, which can be useful for some domains in which the harm needs to be proved [386]. If user experiences shape the design technology, why would not they shape the design of regulations?

11.6 Users as Stakeholders: A Critical Perspective on Educational Interventions

The idea of providing more knowledge to users about the existence of manipulative practices so they are better equipped to protect themselves is extended among scholars and policymakers [341, 215, 287, 399, 420, 434]. Especially, the idea of “data literacy” or “manipulation literacy” has been commonly suggested as a way to address the problem of online manipulation, as explained in Chapter 10. However, this call for users to become more literate about data or manipulative practices embodies problems and risks to consider.

As explained by Pangrazio and Selwyn [352], the call for “literacy” in the digital landscape has become a catchall term lacking a shared definition. Literacy is claimed as the only potential alternative to a regulatory model that treats “digital rights” as property rights: users’ data can be commodified as the base of surveillance capitalism [480]. Thus, as Pangrazio and Sefton-Green [351] follows, educational interventions appear as a saviour when it is unclear whether regulatory interventions can help users by merely giving them control over their online interactions. This is crucial because if users can only protect themselves through the control given by regulators, but vulnerability acts beyond rationality, with impossible trade-offs and under the illusion of agency, users are at a dead end. Regulatory mechanisms may fail because the relationship between users and interfaces is not simply rational. Thus, when users do not manage to protect themselves, although regulators have provided some theoretical mechanisms of protection — e.g., mandatorily asking providers to ask for consent when asking for personal data for tracking mechanisms like cookies [222] —, a victim-blaming logic [300] sometimes accuses users of not doing enough for their own protection.

It is under this idea that users are not well equipped to protect themselves online that literacy arises as a solution related to different domains — e.g., from data protection [353], to digital skills [112, 446], or misinformation [5]. I want to echo Pangrazio and Sefton-Green [351] when they question what we need to educate or what literacy to increase. In the context of this dissertation, it would refer to *manipulative designs literacy*.

While the majority of scholars that suggest literacy and public awareness [230, 412] do it in very generic terms, some scholars have pointed to more specific alternatives like educating users about FOMO —i.e. Fear of Missing Out— design [458], or teaching how to identify manipulative designs on interfaces [230]. There is yet a lot to question about the latter. Should we educate about instances of manipulative designs when they are multimodal and contextual? Or should we educate about a whole system that gives users the appearance of control?

One interesting difference between the participants of the studies presented in Chapters 7, 8 and 9 is the claimed “right to sanctuary”. This “right to sanctuary” coined by Zuboff [480] is the right of not being reached by technology, having one’s own space in which surveillance capitalism cannot reach you. Only participants in Chapter 8 explicitly demanded the right not to use technology. This was unthinkable by participants in the other studies, who tended to normalise the system of manipulative practices they live in. As the present work is not aiming at establishing causation, I cannot claim that the age of participants is the main reason for this difference, but it might add a challenge to develop educational interventions for those generations that are born in a system in which manipulative practices are the norm. I am aware that when Zuboff [480] talks about the risk of surveillance capitalism, she does not necessarily focus on interface design. However, surveillance capitalism is a system in which behaviours, actions, and interactions are used as raw material, transformed into data. UX design is a big contributor to transforming this raw material into data, fuelling the proliferation of manipulative designs. Users do not “have” interactions, users “are” interactions [207]; hence, should we not educate them to understand that their interactions should not be commodified?

Aligned with those interventions that aim to provide knowledge within the interactions, we have suggested in Chapter 10 different strategies that can contribute to generating awareness about harms via representation of harms within the interactions, using the social aspects of harms to share them, translating the temporality of the harm, or designing cumulative experiences. HCI and design knowledge and methods can support educational approaches, and help to translate the drivers of vulnerability into educational material for users. Furthermore, HCI researchers can contribute by creating educational and awareness interventions with more holistic approaches that go beyond steering behaviours in favour of users’ interests — as in the case of bright patterns [162].

11.7 Overall Limitations

Despite the rigour of the study designs, it is important to acknowledge the overall limitations of this work, beyond the limitations reported in each chapter.

This thesis investigated the drivers of vulnerability to manipulative designs by following the third wave of HCI research [123, 149], which values subjective knowledge. The selection of methods thus relied on lived subjective experiences, which do not provide objective and generalizable insights into human behaviour. Relying on lived experiences on a topic highly intertwined with other technical affordances sometimes limits the analysis to qualitative interpretative methods. While not a problem per se, it is important to recognise the boundaries this dissertation establishes when caring about felt experiences. An example of such boundaries is the harms experienced by participants, which often were unveiled by different participants’ testimonies but never directly observed.

Similarly, the studies included in this dissertation mostly rely on episodic methods that rely on memory rather than capturing a live interaction time span using momentary research approaches (as it would be the case for a user test, for instance). These methods entail known

limitations related to the memory of experiences [259, 321, 417]. The contextual elements surrounding the situations described, and subjective experiences of participants might have been different in reality than what users remember and report on. In this trend, it is important to acknowledge the limitation of studying "felt manipulation" as a proxy. While I provided methodological and ontological arguments for using such proxy, this work will be limited in understanding the effects of subtle and hidden mechanisms that sometimes go unnoticed by users. Despite the interpretative nature of some of the analyses that helped to uncover those "hidden" effects (see Chapter 4), it is important to recognise their limitations.

Beyond our own research field, it is fundamental to acknowledge a limitation related to the overall scoping of the topic and the field. Manipulative design research is an important yet still nascent topic in interaction design. However, as most research topics in HCI — in particular, those related to user experiences — it is important to recognise the existence of a wide body of knowledge in other fields. Hence, some of manipulative and deceptive patterns and the associated experiences are also being studied outside our domain. In this dissertation — and despite the willingness to open transdisciplinary conversations and bridge disciplines — the primary focus and scope have been, at times, purportedly limited to HCI (notably in the literature review presented in Chapter 3 or the scoping review of interventions in Chapter 10). This choice is explained by the willingness to consider the state of research in our field and provide empirical insights and interventions spaces, filling some of the gaps in our discipline.

Finally, as I reflected for each chapter on my research identity and positionality, it is also essential to recognise the Western-centric and even Eurocentric approach of this dissertation since it impacts both users and research analysis. The participants involved in our studies are embedded in a system of consumer protection and digital rights (e.g., GDPR) that differs from other countries. For instance, the imposition of consent banners for the use of online trackers — e.g., cookies — on digital services in the European Union gives participants a daily navigation experience different from other parts of the world. Similarly, the same Eurocentric approach has an impact on the analytic lenses used in our studies. Thus, beyond the values I adhere to as a researcher explained in Chapter 4, the set of general values by which European regulations ascribe are also embedded in my analytic approach. Lastly, the lack of familiarity with other legal systems that try to regulate manipulative designs, e.g., the United States, might have limited the interpretation of the results when it comes to collaborating with legal scholars and policymakers.

Chapter 12

Concluding Remarks

This thesis has explored the concept of vulnerability to manipulative designs as a contextual and situated problem to contribute to the community of manipulative design research by providing a more nuanced understanding of vulnerability that can feed design theory and practice. The main objective of this dissertation was to explore the users' drivers of vulnerability in their relationship with manipulative designs. Informed by these insights, the present work additionally aimed to explore design solutions that could help users resist. To do so, the research questions introduced in Chapter 2 have been addressed in this manuscript, and I here conclude with a synthetic overview of how the thesis chapters covered such an investigation.

How do the tensions between persuasive design and manipulation in UX design practice inform vulnerability to manipulative designs? Chapter 5 has been dedicated to address this research question. By working with practitioners, I have elicited two types of tension. First, there is tension between the guiding principles designers use to design to influence users' behaviours in an ethical way and the contextual factors in which their practices take place. Second, there is tension between the guiding principles — autonomy, usability and UI design implementation — and the common UX practices. Through both tensions, I have unveiled how designers can subvert users vulnerabilities even if they do not intend so, giving rise to manipulative design practices.

How can HCI contribute to the conceptualisation of vulnerability to manipulative designs? In Chapter 6, I have provided a multidisciplinary conversation that helps to understand how HCI can contribute to conceptualise vulnerability to manipulative designs. Through theories of perception and interaction, the different ecological factors and systems in which the interaction takes place affect how users interact with manipulative designs, and therefore, it impacts how vulnerability occurs. Thus, beyond the idea of the rational user, I explained how vulnerability would occur when different macro, meso and micro factors drive vulnerability. Additionally it was discussed how HCI can support legal scholars and policymakers in their fight against manipulative designs via empirical evidence, methods that learn from contextual experiences, and tools for designers-legal scholars collaboration — governance models of empowerment and design libraries. I explained how social and contextual differences in users contribute to how users interact with manipulative designs, giving rise to drivers of vulnerability. This theoretical conceptualisation anchors the dissertation and sets the frame for empirically investigating the drivers of vulnerability.

What are the contextual aspects that drive vulnerability to manipulative interfaces? In Chapters 7, 8 and 9, I have studied how different contexts drive vulnerability to manipulative designs with by understanding experiential drivers. Lack of social support, low self-efficacy and the presence of impossible trade-offs in the interface increases the exposure of manipulative designs in users, and also impacts their recovery, making them vulnerable. These experiential drivers are important because they allow us to concretise reasons why users can be vulnerable and are not attached to their socio-demographic conditions, which makes them more actionable to protect users via counter-intervention and better designs.

What are solution spaces to design mitigation strategies for manipulative designs that account

for vulnerability? Chapter 10 has been dedicated to understanding within the experience of manipulation what the different pain points in time, context, and relationships that can lead to vulnerability to manipulative designs. Different solution spaces arise by understanding the temporal experiences of manipulation: leveraging social aspects, supporting goal-setting approaches, learning from intertwined affordances, advancing the experience of harms, designing for resilience and beyond interaction interventions.

By studying what vulnerability means in the context of manipulative designs, this thesis has identified some opportunities for expanding the work on manipulative designs in a way that can protect users. It is not the merit of this thesis to bring approaches and methods that already have a long tradition in HCI and social sciences, but to see their value for the research on manipulative designs and suggest a roadmap of how to apply them. Thus, through the study of vulnerability, I give an overview of the different problems and strengths in the manipulative designs research community that I have found along the way of this dissertation. This has allowed me to provide an overview of the field and suggest future directions for the community.

References

- [1] Jacob Aagaard, Miria Emma Clausen Knudsen, Per Bækgaard, and Kevin Doherty. A game of dark patterns: designing healthy, highly-engaging mobile games. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 1–8, New Orleans LA USA. ACM, Apr. 2022. ISBN: 978-1-4503-9156-6. DOI: [10 . 1145 / 3491101 . 3519837](https://doi.org/10.1145/3491101.3519837).
- [2] Felicia Ackerman. The concept of manipulateness. *Philosophical Perspectives*, 9(AI Connectionism and Philosophical Psychology):335–340, 2021.
- [3] Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, Yang Wang, and Shomir Wilson. Nudges for privacy and security: understanding and assisting users’ choices online. *ACM Computing Surveys*, 50(3):1–41, Oct. 2017. ISSN: 0360-0300, 1557-7341. DOI: [10 . 1145 / 3054926](https://doi.org/10.1145/3054926).
- [4] Gojko Adzic, David Evans, and Nikola Korac. *Fifty Quick Ideas to Improve Your User Stories*. Neuri Consulting LLP, 2014. ISBN: 9780993088100.
- [5] Zhila Aghajari, Eric P. S. Baumer, and Dominic DiFranzo. Reviewing interventions to address misinformation: the need to expand our vision beyond an individualistic focus. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1):1–34, Apr. 2023. ISSN: 2573-0142. DOI: [10 . 1145 / 3579520](https://doi.org/10.1145/3579520).
- [6] Anders Albrechtslund. Ethics and technology design. *Ethics and Information Technology*, 9(1):63–72, Feb. 2007. ISSN: 1388-1957, 1572-8439. DOI: [10 . 1007 / s10676 - 006 - 9129 - 8](https://doi.org/10.1007/s10676-006-9129-8).
- [7] Asam Almohamed, Jinglan Zhang, and Dhaval Vyas. Magic machines for refugees. In *Proceedings of the 3rd ACM SIGCAS Conference on Computing and Sustainable Societies*, 76–86, Ecuador. ACM, June 2020. ISBN: 978-1-4503-7129-2. DOI: [10 . 1145 / 3378393 . 3402256](https://doi.org/10.1145/3378393.3402256).
- [8] Patricia Alves-Oliveira, Maria Luce Lupetti, Michal Luria, Diana Löffler, Mafalda Gamboa, Lea Albaugh, Waki Kamino, Anastasia K. Ostrowski, David Puljiz, Pedro Reynolds-Cuellar, Marcus Scheunemann, Michael Suguitan, and Dan Lockton. Collection of metaphors for human-robot interaction. In *Designing Interactive Systems Conference 2021*, 1366–1379, Virtual Event USA. ACM, June 2021. ISBN: 978-1-4503-8476-6. DOI: [10 . 1145 / 3461778 . 3462060](https://doi.org/10.1145/3461778.3462060).
- [9] Miglena Amirpur and Alexander Benlian. *Buying under Pressure: Purchase Pressure Cues and their Effects on Online Buying Decisions*. Dec. 2015.
- [10] Kristina Andersen. *Making Magic Machines*. Phd Thesis, KTH, Royal Institute of Technology, 2017. URL: <https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1139682&dsid=-2042>.
- [11] Kristina Andersen and Ron Wakkary. The magic machine workshops: making personal design knowledge. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13, Glasgow Scotland Uk. ACM, May 2019. ISBN: 978-1-4503-5970-2. DOI: [10 . 1145 / 3290605 . 3300342](https://doi.org/10.1145/3290605.3300342).
- [12] Jerry Angrave. *The Journey Mapping Playbook: A Practical Guide to Preparing, Facilitating and Unlocking the Value of Customer Journey Mapping*. De Gruyter Business Playbooks Series. De Gruyter, 2020. ISBN: 9783110641110.
- [13] K. Carrie Armel, Aurelie Beaumel, and Antonio Rangel. Biasing simple choices by manipulating relative visual attention. *Judgment and Decision Making*, 3(5):396–403, 2008. ISSN: 1930-2975. DOI: [10 . 1017 / s1930297500000413](https://doi.org/10.1017/s1930297500000413).
- [14] Competition & Markets Authority. *Online Choice Architecture - How digital design can harm competition and consumers - discussion paper*. 2022, page 76.

- [15] Simone Avolicino, Marianna Di Gregorio, Fabio Palomba, Marco Romano, Monica Sebillio, and Giuliana Vitello. *Ai-based emotion recognition to study users' perception of dark patterns*. In *HCI International 2022 - Late Breaking Papers. Design, User Experience and Interaction*. Masaaki Kurosu, Sakae Yamamoto, Hirohiko Mori, Marcelo M. Soares, Elizabeth Rosenzweig, Aaron Marcus, Pei-Luen Patrick Rau, Don Harris, and Wen-Chin Li, editors. Volume 13516. Lecture Notes in Computer Science. Springer International Publishing, Cham, 2022, 185–203. ISBN: 978-3-031-17614-2. DOI: [10.1007/978-3-031-17615-9_13](https://doi.org/10.1007/978-3-031-17615-9_13).
- [16] Pooria Babaei and Julita Vassileva. Drivers and persuasive strategies to influence user intention to learn about manipulative design. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 2421–2431, Rio de Janeiro Brazil. ACM, June 2024. ISBN: 9798400704505. DOI: [10.1145/3630106.3659046](https://doi.org/10.1145/3630106.3659046).
- [17] Stacey Menzel Baker, James W. Gentry, and Terri L. Rittenburg. Building understanding of the domain of consumer vulnerability. *Journal of Macromarketing*, 25(2):128–139, Dec. 2005. ISSN: 0276-1467, 1552-6534. DOI: [10.1177/0276146705280622](https://doi.org/10.1177/0276146705280622).
- [18] Lyndsey L. Bakewell, Konstantina Vasileiou, Kiel S. Long, Mark Atkinson, Helen Rice, Manuela Barreto, Julie Barnett, Michael Wilson, Shaun Lawson, and John Vines. Everything we do, everything we press: data-driven remote performance management in a mobile workplace. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14, Montreal QC Canada. ACM, Apr. 2018. ISBN: 978-1-4503-5620-6. DOI: [10.1145/3173574.3173945](https://doi.org/10.1145/3173574.3173945).
- [19] Jonathan Ball. *The double diamond: a universally accepted depiction of the design process*. In 2019. URL: <https://www.designcouncil.org.uk/our-work/news-opinion/double-diamond-universally-accepted-depiction-design-process/>.
- [20] Nick Ballou, Sebastian Deterding, April Tyack, Elisa D Mekler, Rafael A Calvo, Dorian Peters, Gabriela Villalobos-Zúñiga, and Selen Turkey. Self-determination theory in hci: shaping a research agenda. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 1–6, New Orleans LA USA. ACM, Apr. 2022. ISBN: 978-1-4503-9156-6. DOI: [10.1145/3491101.3503702](https://doi.org/10.1145/3491101.3503702).
- [21] Richard Banfield, C Todd Lombardo, and Trace Wax. *Design sprint: A practical guidebook for building great digital products*. O'Reilly Media, Inc., 2015.
- [22] Jeffrey Bardzell, Shaowen Bardzell, and Erik Stolterman. Reading critical designs: supporting reasoned interpretations of critical design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1951–1960, Toronto Ontario Canada. ACM, Apr. 2014. ISBN: 978-1-4503-2473-1. DOI: [10.1145/2556288.2557137](https://doi.org/10.1145/2556288.2557137).
- [23] Shaowen Bardzell, Jeffrey Bardzell, Jodi Forlizzi, John Zimmerman, and John Antanitis. Critical design and critical theory: the challenge of designing for provocation. In *DIS*, page 10, Newcastle, UK, June 2012.
- [24] Belén Barros Pena, Rachel E Clarke, Lars Erik Holmquist, and John Vines. Circumspect users: older adults as critical adopters and resisters of technology. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–14, Yokohama Japan. ACM, May 2021. ISBN: 978-1-4503-8096-6. DOI: [10.1145/3411764.3445128](https://doi.org/10.1145/3411764.3445128).
- [25] Kristen Barta, Cassidy Pyle, and Nazanin Andalibi. Toward a feminist social media vulnerability taxonomy. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1):1–37, Apr. 2023. ISSN: 2573-0142. DOI: [10.1145/3579533](https://doi.org/10.1145/3579533).
- [26] Christine Barter and Emma Renold. The use of vignettes in qualitative research. URL: <https://sru.soc.surrey.ac.uk/SRU25.html>.
- [27] Belinda Barton, Natalina Zlatevska, and Harmen Oppewal. Scarcity tactics in marketing: a meta-analysis of product scarcity effects on consumer purchase intentions. *Journal of Retailing*, 98(4):741–758, Dec. 2022. ISSN: 0022-4359. DOI: [10.1016/j.jretai.2022.06.003](https://doi.org/10.1016/j.jretai.2022.06.003).
- [28] Eric P. S. Baumer, Mark Blythe, and Theresa Jean Tanenbaum. Evaluating design fiction: the right tool for the job. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, 1901–1913, Eindhoven Netherlands. ACM, July 2020. ISBN: 978-1-4503-6974-9. DOI: [10.1145/3357236.3395464](https://doi.org/10.1145/3357236.3395464).

- [29] Maria Beimborn, Selma Kadi, Nina Köberer, Mara Mühleck, and Mone Spindler. *Focusing on the human: interdisciplinary reflections on ageing and technology*. In *Science Studies*. Emma Domínguez-Rué and Linda Nierling, editors. 1st edition. transcript Verlag, Bielefeld, Germany, 1st edition, Jan. 2016, 311–334. ISBN: 978-3-8376-2957-6. DOI: [10.14361/9783839429570-015](https://doi.org/10.14361/9783839429570-015).
- [30] Daniel Berdichevsky and Erik Neuenschwander. Toward an ethics of persuasive technology. *Communications of the ACM*, 42(5):51–58, May 1999. ISSN: 0001-0782, 1557-7317. DOI: [10.1145/301353.301410](https://doi.org/10.1145/301353.301410).
- [31] Benjamin Maximilian Berens, Heike Dietmann, Chiara Krisam, Oksana Kulyk, and Melanie Volkamer. Cookie disclaimers: impact of design and users' attitude. In *Proceedings of the 17th International Conference on Availability, Reliability and Security*, 1–20, Vienna Austria. ACM, Aug. 2022. ISBN: 978-1-4503-9670-7. DOI: [10.1145/3538969.3539008](https://doi.org/10.1145/3538969.3539008).
- [32] Anders Bergesen, John Gullaksen, Mikael Hanssen, Aleksander Karlsson, and Birgitte Swensson. Dark patterns in cookie consent notices – norway's 50 most visited websites:5.
- [33] Christina Bermeitinger. *Priming*. In *Psychology and Mental Health: Concepts, Methodologies, Tools, and Applications*. IGI Global, 2016, 42–88. ISBN: 978-1-5225-0159-6. DOI: [10.4018/978-1-5225-0159-6.ch003](https://doi.org/10.4018/978-1-5225-0159-6.ch003).
- [34] Carlos Bermejo Fernandez, Dimitris Chatzopoulos, Dimitrios Papadopoulos, and Pan Hui. This website uses nudging: mturk workers' behaviour on cookie consent notices. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–22, Oct. 2021. ISSN: 2573-0142. DOI: [10.1145/3476087](https://doi.org/10.1145/3476087).
- [35] Juan Pablo Bermúdez, Rune Nyrup, Sebastian Deterding, Laura Moradbakhti, Céline Mougénou, Fangzhou You, and Rafael A. Calvo. What is a subliminal technique? an ethical perspective on ai-driven influence. In *2023 IEEE International Symposium on Ethics in Engineering, Science, and Technology (ETHICS)*, 1–10, May 2023. DOI: [10.1109/ethics57328.2023.10155039](https://doi.org/10.1109/ethics57328.2023.10155039).
- [36] Pallabi Bhowmick and Erik Stolterman. Enhancing social connectivity: tangible peer-based check-in systems for isolated older adults. In *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing*, 230–235, Cancun, Quintana Roo Mexico. ACM, Oct. 2023. ISBN: 9798400702006. DOI: [10.1145/3594739.3610764](https://doi.org/10.1145/3594739.3610764).
- [37] Pallabi Bhowmick and Erik Stolterman Bergqvist. Exploring tangible user interface design for social connection among older adults: a preliminary review. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–9, Hamburg Germany. ACM, Apr. 2023. ISBN: 978-1-4503-9422-2. DOI: [10.1145/3544549.3585722](https://doi.org/10.1145/3544549.3585722).
- [38] Nataliia Bielova, Laura Litvine, Anyisia Nguyen, Mariam Chammat, Vincent Toubiana, and Estelle Hary. The effect of design patterns on (present and future) cookie consent decisions, 2024. URL: <https://www.usenix.org/conference/usenixsecurity24/presentation/bielova>.
- [39] J S Blumenthal-Barby. Between reason and coercion: ethically permissible influence in health care and health policy contexts. *Kennedy Institute of Ethics Journal*, 22(4):345–366, 2012.
- [40] Mark Blythe, Kristina Andersen, Rachel Clarke, and Peter Wright. Anti-solutionist strategies: seriously silly design fiction. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4968–4978, San Jose California USA. ACM, May 2016. ISBN: 978-1-4503-3362-7. DOI: [10.1145/2858036.2858482](https://doi.org/10.1145/2858036.2858482).
- [41] European Data Protection Board. *Guidelines 3/2022 on Dark patterns in Social Media Platform Interfaces: How to Recognise and Avoid Them*. Mar. 2022.
- [42] European Data Protection Board. *Guidelines 4/2019 on Article 25 Data Protection by Design and by Default*. Oct. 2020. URL: <https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-42019-article-25-data-protection-design-and-en>.
- [43] European Data Protection Board. Opinion 08/2024 on valid consent in the context of consent or pay models implemented by large online platforms, Apr. 2024. URL: https://www.edpb.europa.eu/system/files/2024-04/edpb_opinion_202408_consentorpay_en.pdf.
- [44] Kirsten Boehner, Janet Vertesi, Phoebe Sengers, and Paul Dourish. How hci interprets the probes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1077–1086, San Jose California USA. ACM, Apr. 2007. ISBN: 978-1-59593-593-9. DOI: [10.1145/1240624.1240789](https://doi.org/10.1145/1240624.1240789).

- [45] Kerstin Bongard-Blanchy, Arianna Rossi, Salvador Rivas, Sophie Doublet, Vincent Koenig, and Gabriele Lenzini. "I am definitely manipulated, even when I am aware of it. it's ridiculous!" - dark patterns from the end-user perspective. In *Designing Interactive Systems Conference 2021*, 763–776, Virtual Event USA. ACM, June 2021. ISBN: 978-1-4503-8476-6. DOI: [10.1145/3461778.3462086](https://doi.org/10.1145/3461778.3462086).
- [46] Ida Borberg, Rene Hougaard, Willard Rafnsson, and Oksana Kulyk. "so i sold my soul": effects of dark patterns in cookie notices on end-user behavior and perceptions. *Workshop on Usable Security and Privacy (USEC)*, 11, May 2022. DOI: [10.14722/usec.2022.23026](https://doi.org/10.14722/usec.2022.23026).
- [47] Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher. Tales from the dark side: privacy dark strategies and privacy dark patterns. *Proceedings on Privacy Enhancing Technologies*, 2016(4):237–254, Oct. 2016. ISSN: 2299-0984. DOI: [10.1515/popets-2016-0038](https://doi.org/10.1515/popets-2016-0038).
- [48] Marietjie Botes. Autonomy and the social dilemma of online manipulative behavior. *AI and Ethics*, Apr. 2022. ISSN: 2730-5953, 2730-5961. DOI: [10.1007/s43681-022-00157-5](https://doi.org/10.1007/s43681-022-00157-5).
- [49] Cennydd Bowles. *Persuasive mechanisms*. In *Future Ethics*. NowNext Press, 2020. ISBN: 978-1-9996019-1-1.
- [50] Virginia Braun and Victoria Clarke. One size fits all? what counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 18(3):328–352, July 2021. ISSN: 1478-0887, 1478-0895. DOI: [10.1080/14780887.2020.1769238](https://doi.org/10.1080/14780887.2020.1769238).
- [51] Virginia Braun and Victoria Clarke. *Successful Qualitative Research. A Practical Guide for Beginners*. Sage, 2013.
- [52] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, Jan. 2006. ISSN: 1478-0887, 1478-0895. DOI: [10.1191/1478088706qp063oa](https://doi.org/10.1191/1478088706qp063oa).
- [53] Bruno G. Breitmeyer, Tony Ro, and Neel S. Singhal. Unconscious color priming occurs at stimulus- not percept-dependent levels of processing. *Psychological Science*, 15(3):198–202, Mar. 2004. ISSN: 0956-7976. DOI: [10.1111/j.0956-7976.2004.01503009.x](https://doi.org/10.1111/j.0956-7976.2004.01503009.x).
- [54] Robin N. Brewer. "if alexa knew the state i was in, it would cry": older adults' perspectives of voice assistants for health. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 1–8, New Orleans LA USA. ACM, Apr. 2022. ISBN: 978-1-4503-9156-6. DOI: [10.1145/3491101.3519642](https://doi.org/10.1145/3491101.3519642).
- [55] Harry Brignull. Dark patterns - types of dark pattern, 2010. URL: <https://www.darkpatterns.org/types-of-dark-pattern>.
- [56] Bringing Dark Patterns to Light Staff Report. Technical report, Federal Trade Commission, Sept. 2022. URL: <https://www.ftc.gov/reports/bringing-dark-patterns-light>.
- [57] Urie Bronfenbrenner. *The ecology of human development: experiments by nature and design*. Harvard University Press, Cambridge, Mass, 1979. ISBN: 978-0-674-22456-8.
- [58] Urie Bronfenbrenner and Pamela A Morris. *The bioecological model of human development*. In *Handbook of Child Psychology*. Wiley, 2007, 793–828. ISBN: 978-0-470-14765-8.
- [59] Ian Brown and Christopher Mardsen. *Regulating Code: Good Governance and Better Regulation in the Information Age*. MIT Press, London, England, 2013.
- [60] Michael Brown, Tim Coughlan, Jesse Blum, Glyn Lawson, Robert Houghton, Richard Mortier, Murray Goulden, and Unna Arunachalam. Tailored scenarios: a low-cost online method to elicit perceptions of home technologies using participant-specific contextual information. *Interacting with Computers*, 27(1):60–71, Jan. 2015. ISSN: 0953-5438, 1873-7951. DOI: [10.1093/iwc/iwu028](https://doi.org/10.1093/iwc/iwu028).
- [61] Michael Bui, Kira Oberschmidt, and Christiane Grünloh. Patient journey value mapping: illustrating values and experiences along the patient journey to support ehealth design. In *Proceedings of Mensch und Computer 2023*, MuC '23, 49–66, New York, NY, USA. Association for Computing Machinery, Sept. 2023. ISBN: 9798400707711. DOI: [10.1145/3603555.3603558](https://doi.org/10.1145/3603555.3603558).
- [62] Brigitte Burgemeestre, Joris Hulstijn, and Yao-Hua Tan. *Rule-based versus principle-based regulatory compliance*. In *Legal Knowledge and Information Systems*. IOS Press, 2009, 37–46. DOI: [10.3233/978-1-60750-082-7-37](https://doi.org/10.3233/978-1-60750-082-7-37).
- [63] Susanne Bødker and Morten Kyng. Participatory design that matters—facing the big issues. *ACM Trans. Comput.-Hum. Interact.*, 25(1):4:1–4:31, Feb. 2018. ISSN: 1073-0516. DOI: [10.1145/3152421](https://doi.org/10.1145/3152421).

- [64] Sofie Krogh Bønlykke, Kamma Amalie Lind Madsen, and Tom Jenkins. Taking the bizarre seriously: dreams as a material for interaction design. In *Designing Interactive Systems Conference*, 699–713, IT University of Copenhagen Denmark. ACM, July 2024. DOI: [10.1145/3643834.3661562](https://doi.org/10.1145/3643834.3661562).
- [65] Jie Cai, Donghee Yvette Wohn, and Guo Freeman. Who purchases and why?: explaining motivations for in-game purchasing in the online survival game fortnite. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 391–396, Barcelona Spain. ACM, 2019. ISBN: 978-1-4503-6688-5. DOI: [10.1145/3311350.3347196](https://doi.org/10.1145/3311350.3347196).
- [66] Gabriela Cajamarca, Valeria Herskovic, Andrés Lucero, and Angeles Aldunate. A co-design approach to explore health data representation for older adults in chile and ecuador. In *Designing Interactive Systems Conference*, 1802–1817, Virtual Event Australia. ACM, June 2022. ISBN: 978-1-4503-9358-4. DOI: [10.1145/3532106.3533558](https://doi.org/10.1145/3532106.3533558).
- [67] Cormac Callanan. *Evolution, implementation and practice of internet self-regulation, co-regulation and public-private collaboration*. In *Self- and Co-regulation in Cybercrime, Cybersecurity and National Security*. Springer-Briefs in Cybersecurity. Springer International Publishing, Cham, 2015, 43–100. ISBN: 978-3-319-16446-5. DOI: [10.1007/978-3-319-16447-2_2](https://doi.org/10.1007/978-3-319-16447-2_2).
- [68] Ryan Calo. *Privacy, vulnerability, and affordance*. In *The Cambridge Handbook of Consumer Privacy*. Evan Selinger, Jules Polonetsky, and Omer Tene, editors. 1st edition. Cambridge University Press, 1st edition, Mar. 2018, 198–206. ISBN: 978-1-316-83196-0. DOI: [10.1017/9781316831960.011](https://doi.org/10.1017/9781316831960.011).
- [69] Nadia Campo Woytuk, Linette Nilsson, and Mingxing Liu. Your period rules: design implications for period-positive technologies. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–6, Glasgow Scotland Uk. ACM, May 2019. ISBN: 978-1-4503-5971-9. DOI: [10.1145/3290607.3312888](https://doi.org/10.1145/3290607.3312888).
- [70] Evan Caragay, Katherine Xiong, Jonathan Zong, and Daniel Jackson. Beyond dark patterns: a concept-based framework for ethical software design. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–16, Honolulu HI USA. ACM, May 2024. ISBN: 9798400703300. DOI: [10.1145/3613904.3642781](https://doi.org/10.1145/3613904.3642781).
- [71] Rachele Carli and Davide Calvaresi. Reinterpreting vulnerability to tackle deception in principles-based xai for human-computer interaction. In *Explainable and Transparent AI and Multi-Agent Systems: 5th International Workshop, EXTRAAMAS 2023, London, UK, May 29, 2023, Revised Selected Papers*, 249–269, Berlin, Heidelberg. Springer-Verlag, Sept. 2023. ISBN: 978-3-031-40877-9. DOI: [10.1007/978-3-031-40878-6_14](https://doi.org/10.1007/978-3-031-40878-6_14).
- [72] Ana Maria Carrillo and Roxana Widmer-Iliescu. *Ageing in a digital world – from vulnerable to valuable*. ITU Publications. 2021, page 50. URL: <http://handle.itu.int/11.1002/pub/818a937f-en>.
- [73] Ian Carter. *Positive and negative liberty*. In *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta, editor. spring 2022 edition. Metaphysics Research Lab, Stanford University, spring 2022 edition, 2022. URL: <https://plato.stanford.edu/archives/spr2022/entries/liberty-positive-negative/>.
- [74] Annette Cerulli-Harms, Marlene Münsch, Christian Thorun, Frithjof Michalsen, and Pierre Hausemer. Loot boxes in online games and their effect on consumers, in particular young consumers. In page 56. Policy Department for Economic, Scientific and Quality of Life Policies. European Parliament., 2020. URL: [https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU\(2020\)652727](https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2020)652727).
- [75] Diana Yian Lian Chan, Shaun Wen Huey Lee, and Pei-Lee Teh. Factors influencing technology use among low-income older adults: a systematic review. *Heliyon*, 9(9):e20111, Sept. 2023. ISSN: 24058440. DOI: [10.1016/j.heliyon.2023.e20111](https://doi.org/10.1016/j.heliyon.2023.e20111).
- [76] Weichen Joe Chang, Katie Seaborn, and Andrew A. Adams. Theorizing deception: a scoping review of theory in research on dark patterns and deceptive design. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems*, Chi Ea '24, 1–7, New York, NY, USA. Association for Computing Machinery, May 2024. ISBN: 9798400703317. DOI: [10.1145/3613905.3650997](https://doi.org/10.1145/3613905.3650997).
- [77] Tanya L. Chartrand, Rick B. van Baaren, and John A. Bargh. Linking automatic evaluation to mood and information processing style: consequences for experienced affect, impression formation, and stereotyping. *Journal of Experimental Psychology: General*, 135(1):70–77, 2006. ISSN: 1939-2222. DOI: [10.1037/0096-3445.135.1.70](https://doi.org/10.1037/0096-3445.135.1.70).

- [78] Akash Chaudhary, Jaivrat Saroha, Kyzyl Monteiro, Angus G. Forbes, and Aman Parnami. "Are you still watching?": exploring unintended user behaviors and dark patterns on video streaming platforms. In *Designing Interactive Systems Conference*, 776–791, Virtual Event Australia. ACM, June 2022. ISBN: 978-1-4503-9358-4. DOI: [10.1145/3532106.3533562](https://doi.org/10.1145/3532106.3533562). URL: <https://dl.acm.org/doi/10.1145/3532106.3533562>.
- [79] Jengchung Victor Chen, Sirapattra Ruangsri, Quang-An Ha, and Andree E. Widjaja. An experimental study of consumers' impulse buying behaviour in augmented reality mobile shopping apps. *Behaviour & Information Technology*:1–22, 2021. ISSN: 0144-929x, 1362-3001. DOI: [10.1080/0144929x.2021.1987523](https://doi.org/10.1080/0144929x.2021.1987523).
- [80] Jieshan Chen, Jiamou Sun, Sidong Feng, Zhenchang Xing, Qinghua Lu, Xiwei Xu, and Chunyang Chen. Unveiling the tricks: automated detection of dark patterns in mobile applications. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 1–20, San Francisco CA USA. ACM, Oct. 2023. ISBN: 9798400701320. DOI: [10.1145/3586183.3606783](https://doi.org/10.1145/3586183.3606783).
- [81] Anna Cherenshchykova and Andrew D. Miller. Sociotechnical design opportunities for pervasive family sleep technologies. In *Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '20, 11–20, New York, NY, USA. Association for Computing Machinery, Feb. 2021. ISBN: 978-1-4503-7532-0. DOI: [10.1145/3421937.3421979](https://doi.org/10.1145/3421937.3421979).
- [82] Shruthi Sai Chivukula, Jason Brier, and Colin M. Gray. Dark intentions or persuasion?: UX designers' activation of stakeholder and user values. In *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems*, 87–91, Hong Kong China. ACM, May 2018. ISBN: 978-1-4503-5631-2. DOI: [10.1145/3197391.3205417](https://doi.org/10.1145/3197391.3205417).
- [83] Shruthi Sai Chivukula, Colin M. Gray, and Jason A. Brier. Analyzing value discovery in design decisions through ethnography. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12, Glasgow Scotland Uk. ACM, May 2019. ISBN: 978-1-4503-5970-2. DOI: [10.1145/3290605.3300307](https://doi.org/10.1145/3290605.3300307).
- [84] Shruthi Sai Chivukula, Aiza Hasib, Ziqing Li, Jingle Chen, and Colin M. Gray. Identity claims that underlie ethical awareness and action. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–13, Yokohama Japan. ACM, May 2021. ISBN: 978-1-4503-8096-6. DOI: [10.1145/3411764.3445375](https://doi.org/10.1145/3411764.3445375).
- [85] Shruthi Sai Chivukula, Ziqing Li, Anne C. Pivonka, Jingning Chen, and Colin M. Gray. Surveying the landscape of ethics-focused design methods. *arXiv:2102.08909 [cs]*, Feb. 2021. URL: <http://arxiv.org/abs/2102.08909>.
- [86] Shruthi Sai Chivukula, Chris Watkins, Lucca McKay, and Colin M. Gray. "nothing comes before profit": asshole design in the wild. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–6, Glasgow Scotland Uk. ACM, May 2019. ISBN: 978-1-4503-5971-9. DOI: [10.1145/3290607.3312863](https://doi.org/10.1145/3290607.3312863).
- [87] Shruthi Sai Chivukula, Chris Rhys Watkins, Rhea Manocha, Jingle Chen, and Colin M. Gray. Dimensions of ux practice that shape ethical awareness. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13, Honolulu HI USA. ACM, Apr. 2020. ISBN: 978-1-4503-6708-0. DOI: [10.1145/3313831.3376459](https://doi.org/10.1145/3313831.3376459).
- [88] Ishita Chordia, Lena-Phuong Tran, Tala June Tayebi, Emily Parrish, Sheena Erete, Jason Yip, and Alexis Hiniker. Deceptive design patterns in safety technologies: a case study of the citizen app. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Chi '23, 1–18, New York, NY, USA. Association for Computing Machinery, Apr. 2023. ISBN: 978-1-4503-9421-5. DOI: [10.1145/3544548.3581258](https://doi.org/10.1145/3544548.3581258).
- [89] Robert B. Cialdini. *Influence, New and Expanded: The Psychology of Persuasion*. HarperCollins, New York, 2021.
- [90] Danielle Keats Citron and Daniel J. Solove. Privacy harms. (3782222), Feb. 2021. DOI: [10.2139/ssrn.3782222](https://doi.org/10.2139/ssrn.3782222).
- [91] James M. Clarke, Maryam Mehrnezhad, and Ehsan Toreini. Invisible, unreadable, and inaudible cookie notices: an evaluation of cookie notices for users with visual impairments. *ACM Trans. Access. Comput.*, 17(1):1:1–1:39, Mar. 2024. ISSN: 1936-7228. DOI: [10.1145/3641281](https://doi.org/10.1145/3641281).
- [92] European Commission. Digital fairness - fitness check on eu consumer law, Text, June 2022. URL: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13413-Digital-fairness-fitness-check-on-EU-consumer-law_en.

- [93] European Commission. *Guidance on the interpretation and application of Directive 2005/29/EC of the European Parliament and of the Council concerning unfair business-to-consumer commercial practices in the internal market*, number 2021/C 526/01. Dec. 2021. URL: <https://op.europa.eu/en/publication-detail/-/publication/c608fff7-687a-11ec-9136-01aa75ed71a1/language-en>.
- [94] Silvia De Conca. The present looks nothing like the jetsons: deceptive design in virtual assistants and the protection of the rights of users. *Computer Law & Security Review*, 51:105866, Nov. 2023. ISSN: 0267-3649. DOI: [10.1016/j.clsr.2023.105866](https://doi.org/10.1016/j.clsr.2023.105866).
- [95] Gregory Conti and Edward Sobiesk. Malicious interface design: exploiting the user. In *Proceedings of the 19th international conference on World wide web - WWW '10*, page 271, Raleigh, North Carolina, USA. ACM Press, 2010. ISBN: 978-1-60558-799-8. DOI: [10.1145/1772690.1772719](https://doi.org/10.1145/1772690.1772719).
- [96] Heith Copes, Whitney Tchoula, Fiona Brookman, and Jared Ragland. Photo-elicitation interviews with vulnerable populations: practical and ethical considerations. *Deviant Behavior*, 39(4):475–494, Apr. 2018. ISSN: 0163-9625, 1521-0456. DOI: [10.1080/01639625.2017.1407109](https://doi.org/10.1080/01639625.2017.1407109).
- [97] Mayara Costa Figueiredo and Yunan Chen. Health data in fertility care: an ecological perspective. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–17, Yokohama Japan. ACM, May 2021. ISBN: 978-1-4503-8096-6. DOI: [10.1145/3411764.3445189](https://doi.org/10.1145/3411764.3445189).
- [98] Sasha Costanza-Chock. Design justice: towards an intersectional feminist framework for design theory and practice. In *Design Research Society Conference, 2018*, June 2018. DOI: [10.21606/drs.2018.679](https://doi.org/10.21606/drs.2018.679).
- [99] Francis Joseph Costello, Jinho Yun, and Kun Chang Lee. Digital dark nudge: an exploration of when digital nudges unethically depart. In 2022. DOI: [10.24251/HICSS.2022.531](https://doi.org/10.24251/HICSS.2022.531). URL: <http://hdl.handle.net/10125/79868>.
- [100] Antoine Courmont. Le plaignant type? un homme, diplômé et cadre, Feb. 2022. URL: <https://linc.cnil.fr/le-plaignant-type-un-homme-diplome-et-cadre>.
- [101] Anna L. Cox, Sandy J.J. Gould, Marta E. Cecchinato, Ioanna Iacovides, and Ian Renfree. Design frictions for mindful interactions: the case for microboundaries. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 1389–1397, San Jose California USA. ACM, May 2016. ISBN: 978-1-4503-4082-3. DOI: [10.1145/2851581.2892410](https://doi.org/10.1145/2851581.2892410).
- [102] Caroline Crump, Stephen Walenchok, Chris Johnson, Joseph Pauszek, and Douglas Young. Stress, operator error, and transportation accidents. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 65(1):1453–1458, Sept. 2021. ISSN: 1071-1813. DOI: [10.1177/1071181321651171](https://doi.org/10.1177/1071181321651171).
- [103] Andrea Curley, Dymrna O'Sullivan, Damian Gordon, Brendan Tierney, and Ioannis Stavrakakis. The design of a framework for the detection of web-based dark patterns. In *ICDS 2021: The 15th International Conference on Digital Society*, page 8, Nice, France (Online), July 2021.
- [104] Claudia Daudén Roquet and Corina Sas. Body matters: exploration of the human body as a resource for the design of technologies for meditation. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, 533–546, Eindhoven Netherlands. ACM, July 2020. ISBN: 978-1-4503-6974-9. DOI: [10.1145/3357236.3395499](https://doi.org/10.1145/3357236.3395499).
- [105] Janet Davis. Design methods for ethical persuasive computing. In *Proceedings of the 4th International Conference on Persuasive Technology - Persuasive '09*, page 1, Claremont, California. ACM Press, 2009. ISBN: 978-1-60558-376-1. DOI: [10.1145/1541948.1541957](https://doi.org/10.1145/1541948.1541957).
- [106] Martin Johannes Dechant, Max Valentin Birk, and Kathrin Gerling. Vulnerability and play: exploring the mundane ways in which games might harm players. In *Companion Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 225–227, Stratford ON Canada. ACM, Oct. 2023. ISBN: 9798400700293. DOI: [10.1145/3573382.3616048](https://doi.org/10.1145/3573382.3616048).
- [107] Edward L Deci and Richard M Ryan. *Handbook of self-determination research*. University Rochester Press, 2004.
- [108] Laura Dempsey, Maura Dowling, Philip Larkin, and Kathy Murphy. Sensitive interviewing in qualitative research: sensitive interviewing. *Research in Nursing & Health*, 39(6):480–490, Dec. 2016. ISSN: 01606891. DOI: [10.1002/nur.21743](https://doi.org/10.1002/nur.21743).

- [109] Alan R. Dennis, Lingyao (IVY) Yuan, Xuan Feng, Eric Webb, and Christine J. Hsieh. Digital nudging: numeric and semantic priming in e-commerce. *Journal of Management Information Systems*, 37(1):39–65, Jan. 2020. ISSN: 0742-1222. DOI: [10.1080/07421222.2019.1705505](https://doi.org/10.1080/07421222.2019.1705505).
- [110] Design patterns – données & design par linc. en-US. URL: <https://design.cnil.fr/en/design-patterns/>.
- [111] Linda Di Geronimo, Larissa Braz, Enrico Fregnan, Fabio Palomba, and Alberto Bacchelli. UI dark patterns and where to find them: a study on mobile applications and user perception. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14, Honolulu HI USA. ACM, Apr. 2020. ISBN: 978-1-4503-6708-0. DOI: [10.1145/3313831.3376600](https://doi.org/10.1145/3313831.3376600).
- [112] Paul DiMaggio and Eszter Hargittai. From the “Digital Divide” to “Digital Inequality”: studying internet use as penetration increases. *Working Paper Series*. Center for Arts and Cultural Policy Studies, 15:25, 2001.
- [113] Daniella DiPaola and Ryan Calo. Socio-digital vulnerability, Jan. 2024. URL: <http://dx.doi.org/10.2139/ssrn.4686874>.
- [114] Verena Distler. The influence of context on response to spear-phishing attacks: an in-situ deception study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–18, Hamburg Germany. ACM, Apr. 2023. ISBN: 978-1-4503-9421-5. DOI: [10.1145/3544548.3581170](https://doi.org/10.1145/3544548.3581170).
- [115] Verena Distler, Matthias Fassl, Hana Habib, Katharina Krombholz, Gabriele Lenzini, Carine Lallemand, Lorrie Faith Cranor, and Vincent Koenig. A systematic literature review of empirical methods and risk representation in usable privacy and security research. *ACM Transactions on Computer-Human Interaction*, 28(6):1–50, Dec. 2021. ISSN: 1073-0516, 1557-7325. DOI: [10.1145/3469845](https://doi.org/10.1145/3469845).
- [116] Verena Distler, Matthias Fassl, Hana Habib, Katharina Krombholz, Gabriele Lenzini, Carine Lallemand, Vincent Koenig, and Lorrie Faith Cranor. *Empirical research methods in usable privacy and security*. In *Human Factors in Privacy Research*. Nina Gerber, Alina Stöver, and Karola Marky, editors. Springer International Publishing, Cham, 2023, 29–53. ISBN: 978-3-031-28642-1. DOI: [10.1007/978-3-031-28643-8_3](https://doi.org/10.1007/978-3-031-28643-8_3).
- [117] Verena Distler, Carine Lallemand, and Vincent Koenig. Making encryption feel secure: investigating how descriptions of encryption impact perceived security. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 220–229, Genoa, Italy. Ieee, Sept. 2020. ISBN: 978-1-72818-597-2. DOI: [10.1109/EuroSPW51379.2020.00037](https://doi.org/10.1109/EuroSPW51379.2020.00037).
- [118] Verena Distler, Gabriele Lenzini, Carine Lallemand, and Vincent Koenig. The framework of security-enhancing friction: how ux can help users behave more securely. In *New Security Paradigms Workshop 2020*, 45–58, Online USA. ACM, Oct. 2020. ISBN: 978-1-4503-8995-2. DOI: [10.1145/3442167.3442173](https://doi.org/10.1145/3442167.3442173).
- [119] Nazarov Dmitry and Baimukhambetov Yerkebulan. Clustering of dark patterns in the user interfaces of websites and online trading portals (E-Commerce). *Science in China, Series A: Mathematics*, 10(18):3219, Sept. 2022. ISSN: 2227-7390, 2227-7390. DOI: [10.3390/math10183219](https://doi.org/10.3390/math10183219).
- [120] M. Dodel, F. Aguirre, and Consulting teams. Disto questionnaire Uruguay adults in Spanish: cuestionario wip/disto uruguay adultos en espanol. a DISTO project survey. 2017. URL: <https://www.lse.ac.uk/media-and-communications/research/research-projects/disto/surveys..>
- [121] Laura Dove, Stephen Reinach, and Irwin Kwan. Lightweight journey mapping: the integration of marketing and user experience through customer driven narratives. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, Chi Ea '16, 880–888, New York, NY, USA. Association for Computing Machinery, May 2016. ISBN: 978-1-4503-4082-3. DOI: [10.1145/2851581.2851608](https://doi.org/10.1145/2851581.2851608).
- [122] M. Dreier, K. Wölfling, E. Duven, S. Giral, M.E. Beutel, and K.W. Müller. Free-to-play: about addicted whales, at risk dolphins and healthy minnows. monetarization design and internet gaming disorder. *Addictive Behaviors*, 64:328–333, Jan. 2017. ISSN: 03064603. DOI: [10.1016/j.addbeh.2016.03.008](https://doi.org/10.1016/j.addbeh.2016.03.008).
- [123] Emanuel Felipe Duarte and M. Cecilia C. Baranauskas. Revisiting the three HCI waves: A preliminary discussion on philosophy of science and research paradigms. In *Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems*, 1–4, São Paulo Brazil. ACM, Oct. 2016. DOI: [10.1145/3033701.3033740](https://doi.org/10.1145/3033701.3033740).

- [124] Mateusz Dubiel, Anastasia Sergeeva, and Luis A. Leiva. Impact of voice fidelity on decision making: a potential dark pattern? In *Proceedings of the 29th International Conference on Intelligent User Interfaces*, 181–194, Greenville SC USA. ACM, Mar. 2024. ISBN: 9798400705083. DOI: [10.1145/3640543.3645202](https://doi.org/10.1145/3640543.3645202).
- [125] Martin Ebers, Veronica R. S. Hoch, Frank Rosenkranz, Hannah Ruschemeier, and Björn Steinrötter. The european commission’s proposal for an artificial intelligence act—a critical assessment by members of the robotics and ai law society (rails). *J. 4(44)*:589–603, Dec. 2021. ISSN: 2571-8800. DOI: [10.3390/j4040043](https://doi.org/10.3390/j4040043).
- [126] London Economics, VVA Consulting, and Ipsos Mori consortium. *Consumer vulnerability across key markets in the European Union. Final report*. Jan. 2016. URL: https://commission.europa.eu/system/files/2018-04/consumers-approved-report_en.pdf.
- [127] Zohar Efroni. The digital services act: risk-based regulation of online platforms. *Internet Policy Review*, Nov. 2021. URL: <https://policyreview.info/articles/news/digital-services-act-risk-based-regulation-online-platforms/1606>.
- [128] Chloe Egtebas, Gudrun Klinker, Susanne Boll, and Marion Koelle. Co-speculating on dark scenarios and unintended consequences of a ubiquitous(ly) augmented reality. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, Dis ’23, 2392–2407, New York, NY, USA. Association for Computing Machinery, July 2023. ISBN: 978-1-4503-9893-0. DOI: [10.1145/3563657.3596073](https://doi.org/10.1145/3563657.3596073).
- [129] Upol Ehsan and Mark O. Riedl. Explainability pitfalls: beyond dark patterns in explainable AI. *Patterns*, 5(6):100971, June 2024. DOI: [10.1016/j.patter.2024.100971](https://doi.org/10.1016/j.patter.2024.100971).
- [130] Martin Eisend and Farid Tarrahi. Persuasion knowledge in the marketplace: a meta-analysis. *Journal of Consumer Psychology*, 32(1):3–22, Jan. 2022. ISSN: 1057-7408, 1532-7663. DOI: [10.1002/jcpy.1258](https://doi.org/10.1002/jcpy.1258).
- [131] Mohamed Elgendi, Parmod Kumar, Skye Barbic, Newton Howard, Derek Abbott, and Andrzej Cichocki. Subliminal priming—state of the art and future perspectives. *Behavioral Sciences*, 8(6):54, May 2018. ISSN: 2076-328X. DOI: [10.3390/bs8060054](https://doi.org/10.3390/bs8060054).
- [132] Evidence review of Online Choice Architecture and consumer and competition harm. Technical report, 2022. URL: <https://www.gov.uk/government/publications/online-choice-architecture-how-digital-design-can-harm-competition-and-consumers/evidence-review-of-online-choice-architecture-and-consumer-and-competition-harm>. Accessed: 2022-4-13.
- [133] World Leaders in Research-Based User Experience. A 100-year view of user experience (by jakob nielsen). URL: <https://www.nngroup.com/articles/100-years-ux/>.
- [134] Rebecca Eynon and Anne Geniets. The digital skills paradox: how do digitally excluded youth develop skills to use the internet? *Learning, Media and Technology*, 41(3):463–479, July 2016. ISSN: 1743-9884, 1743-9892. DOI: [10.1080/17439884.2014.1002845](https://doi.org/10.1080/17439884.2014.1002845).
- [135] Madison Fansher, Shruthi Sai Chivukula, and Colin M. Gray. #Darkpatterns: UX practitioner conversations about ethical design. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–6, Montreal QC Canada. ACM, Apr. 2018. ISBN: 978-1-4503-5621-3. DOI: [10.1145/3170427.3188553](https://doi.org/10.1145/3170427.3188553).
- [136] Martha Fineman. The vulnerable subject: anchoring equality in the human condition. In *Transcending the Boundaries of Law*, 177–191, 0th edition. Routledge-Cavendish, July 2010. ISBN: 978-0-203-84853-1. DOI: [10.4324/9780203848531-26](https://doi.org/10.4324/9780203848531-26).
- [137] Martha Albertson Fineman. Beyond identities: the limits of an antidiscrimination approach to equality. *Boston University Law Review*, 92.
- [138] Martha Albertson Fineman. Vulnerability and social justice. *SSRN Electronic Journal*, 2019. ISSN: 1556-5068. DOI: [10.2139/ssrn.3352825](https://doi.org/10.2139/ssrn.3352825).
- [139] Dan Fitton, Beth T. Bell, and Janet C. Read. *Integrating dark patterns into the 4cs of online risk in the context of young people and mobile gaming apps*. In *Human-Computer Interaction – INTERACT 2021*. Carmelo Ardito, Rosa Lanzilotti, Alessio Malizia, Helen Petrie, Antonio Piccinno, Giuseppe Desolda, and Kori Inkpen, editors. Volume 12935. Lecture Notes in Computer Science. Springer International Publishing, Cham, 2021, 701–711. ISBN: 978-3-030-85609-0. DOI: [10.1007/978-3-030-85610-6_40](https://doi.org/10.1007/978-3-030-85610-6_40).

- [140] Dan Fitton and Janet C. Read. Creating a framework to support the critical consideration of dark design aspects in free-to-play apps. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*, 407–418, Boise ID USA. ACM, June 2019. ISBN: 978-1-4503-6690-8. DOI: [10.1145/3311927.3323136](https://doi.org/10.1145/3311927.3323136).
- [141] BJ Fogg. *Computers as persuasive tools*. In *Persuasive Technology*. Stanford University, 2003, 31–59.
- [142] BJ Fogg. *The ethics of persuasive technology*. In *Persuasive Technology*. Morgan Kaufmann, Stanford, 2009, 211–240.
- [143] ACM. Association for Computing Machinery. Words matter: alternatives for charged terminology in the computing profession, 2023. URL: <https://www.acm.org/diversity-inclusion/words-matter>.
- [144] International Organization for Standardization. Iso 22458:2022 consumer vulnerability — requirements and guidelines for the design and delivery of inclusive service, 2022. URL: <https://www.iso.org/standard/73261.html>.
- [145] International Organization for Standardization. Iso 31700-1:2023 consumer protection — privacy by design for consumer goods and services, 2023. URL: <https://www.iso.org/standard/84977.html>.
- [146] ForbrukerRadet. *Insert Coin. How the Gaming Industry Exploits Consumers Using Loot Boxes*. May 2022. URL: <https://storage02.forbrukerradet.no/media/2022/05/2022-05-31-insert-coin-publish.pdf>.
- [147] Jolien C. Francken, Simon van Gaal, and Floris P. de Lange. Immediate and long-term priming effects are independent of prime awareness. *eng. Consciousness and Cognition*, 20(4):1793–1800, Dec. 2011. ISSN: 1090-2376. DOI: [10.1016/j.concog.2011.04.005](https://doi.org/10.1016/j.concog.2011.04.005).
- [148] Thomas Franke, Christiane Attig, and Daniel Wessel. A personal resource for technology interaction: development and validation of the affinity for technology interaction (ati) scale. *International Journal of Human-Computer Interaction*, 35(6):456–467, Apr. 2019. ISSN: 1044-7318, 1532-7590. DOI: [10.1080/10447318.2018.1456150](https://doi.org/10.1080/10447318.2018.1456150).
- [149] Christopher Frauenberger. Entanglement hci the next wave? *ACM Transactions on Computer-Human Interaction*, 27(1):1–27, Feb. 2020. ISSN: 1073-0516, 1557-7325. DOI: [10.1145/3364998](https://doi.org/10.1145/3364998).
- [150] Guo Freeman, Karen Wu, Nicholas Nower, and Donghee Yvette Wohn. Pay to win or pay to cheat: how players of competitive online games perceive fairness of in-game purchases. *Proceedings of the ACM on Human-Computer Interaction*, 6(Chi Play):1–24, 2022. ISSN: 2573-0142. DOI: [10.1145/3549510](https://doi.org/10.1145/3549510).
- [151] Jeremy Freese, Salvador Rivas, and Eszter Hargittai. Cognitive ability and internet use among older adults. *Poetics*, 34(4–5):236–249, Aug. 2006. ISSN: 0304422X. DOI: [10.1016/j.poetic.2006.05.008](https://doi.org/10.1016/j.poetic.2006.05.008).
- [152] Paulo Freire. *Chapter 1*. In *Pedagogy of the oppressed*. Harmondworth: Penguin, 1990, 20–44.
- [153] Batya Friedman, Peter H Kahn, and Alan Borning. Value sensitive design: theory and methods:25, 2003.
- [154] Alisa Frik, Julia Bernd, Noura Alomar, and Serge Egelman. A qualitative model of older adults' contextual decision-making about information sharing:62, 2020.
- [155] Verena Fuchsberger and Christopher Frauenberger. Doing responsibilities in entangled worlds. *Human-Computer Interaction*:1–24, Oct. 2023. DOI: [10.1080/07370024.2023.2269934](https://doi.org/10.1080/07370024.2023.2269934).
- [156] Natalie Garrett, Nathan Beard, and Casey Fiesler. More than “if time allows”: the role of ethics in ai education. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 272–278, New York NY USA. ACM, Feb. 2020. ISBN: 978-1-4503-7110-0. DOI: [10.1145/3375627.3375868](https://doi.org/10.1145/3375627.3375868). URL: <https://dl.acm.org/doi/10.1145/3375627.3375868>.
- [157] Cally Gatehouse, Matthew Wood, Jo Briggs, James Pickles, and Shaun Lawson. Troubling vulnerability: designing with lgbt young people's ambivalence towards hate crime reporting. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13, Montreal QC Canada. ACM, Apr. 2018. ISBN: 978-1-4503-5620-6. DOI: [10.1145/3173574.3173683](https://doi.org/10.1145/3173574.3173683).
- [158] Bill Gaver, Tony Dunne, and Elena Pacenti. Design: cultural probes. *Interactions*, 6(1):21–29, Jan. 1999. ISSN: 1072-5520, 1558-3449. DOI: [10.1145/291224.291235](https://doi.org/10.1145/291224.291235).

- [159] Erin Gibson, Mark D. Griffiths, Filipa Calado, and Andrew Harris. The relationship between videogame micro-transactions and problem gaming and gambling: a systematic review. *Computers in Human Behavior*, 131, June 2022. DOI: [10.1016/j.chb.2022.107219](https://doi.org/10.1016/j.chb.2022.107219).
- [160] James J. Gibson. *The theory of affordances*. In *The Ecological Approach to Visual Perception*. Lawrence Erlbaum Associates, 1986, 127–137.
- [161] Brendan Gough and Anna Madill. Subjectivity in psychological science: from problem to prospect. *Psychological Methods*, 17(3):374–384, Sept. 2012. ISSN: 1939-1463, 1082-989x. DOI: [10.1037/a0029313](https://doi.org/10.1037/a0029313).
- [162] Paul Graßl, Hanna Schraffenberger, Frederik Zuiderveen Borgesius, and Moniek Buijzen. Dark and bright patterns in cookie consent requests. 3(1):38, 2021.
- [163] Colin M. Gray, Jingle Chen, Shruthi Sai Chivukula, and Liyang Qu. End user accounts of dark patterns as felt manipulation. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–25, Oct. 2021. ISSN: 2573-0142. DOI: [10.1145/3479516](https://doi.org/10.1145/3479516).
- [164] Colin M. Gray and Shruthi Sai Chivukula. Ethical mediation in UX practice. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–11, Glasgow Scotland Uk. ACM, May 2019. ISBN: 978-1-4503-5970-2. DOI: [10.1145/3290605.3300408](https://doi.org/10.1145/3290605.3300408).
- [165] Colin M. Gray, Shruthi Sai Chivukula, Thomas V Carlock, Ziqing Li, and Ja-Nae Duane. Scaffolding ethics-focused methods for practice resonance. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, 2375–2391, Pittsburgh PA USA. ACM, July 2023. DOI: [10.1145/3563657.3596111](https://doi.org/10.1145/3563657.3596111).
- [166] Colin M. Gray, Shruthi Sai Chivukula, and Ahreum Lee. What kind of work do “asshole designers” create? describing properties of ethical concern on reddit. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, 61–73, Eindhoven Netherlands. ACM, July 2020. ISBN: 978-1-4503-6974-9. DOI: [10.1145/3357236.3395486](https://doi.org/10.1145/3357236.3395486).
- [167] Colin M. Gray, Shruthi Sai Chivukula, Cassandra Melkey, and Rhea Manocha. Understanding “dark” design roles in computing education. In *Proceedings of the 17th ACM Conference on International Computing Education Research*, 225–238, Virtual Event USA. ACM, Aug. 2021. ISBN: 978-1-4503-8326-4. DOI: [10.1145/3446871.3469754](https://doi.org/10.1145/3446871.3469754).
- [168] Colin M. Gray, Johanna T. Gunawan, René Schäfer, Nataliia Bielova, Lorena Sanchez Chamorro, Katie Seaborn, Thomas Mildner, and Hauke Sandhaus. Mobilizing research and regulatory action on dark patterns and deceptive design practices. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 1–6, Honolulu HI USA. ACM, May 2024. ISBN: 9798400703317. DOI: [10.1145/3613905.3636310](https://doi.org/10.1145/3613905.3636310).
- [169] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. The dark (patterns) side of UX design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14, Montreal QC Canada. ACM, Apr. 2018. ISBN: 978-1-4503-5620-6. DOI: [10.1145/3173574.3174108](https://doi.org/10.1145/3173574.3174108).
- [170] Colin M. Gray, Thomas Mildner, and Nataliia Bielova. Temporal analysis of dark patterns: a case study of a user’s odyssey to conquer prime membership cancellation through the “iliad flow”. (arXiv:2309.09635), Sept. 2023. URL: <http://arxiv.org/abs/2309.09635>. arXiv:2309.09635 [cs].
- [171] Colin M. Gray, Lorena Sánchez Chamorro, Ike Obi, and Ja-Nae Duane. Mapping the landscape of dark patterns scholarship: a systematic literature review. In *Designing Interactive Systems Conference*, 188–193, Pittsburgh PA USA. ACM, July 2023. ISBN: 978-1-4503-9898-5. DOI: [10.1145/3563703.3596635](https://doi.org/10.1145/3563703.3596635).
- [172] Colin M. Gray, Cristiana Santos, and Nataliia Bielova. Towards a preliminary ontology of dark patterns knowledge. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–9, Hamburg Germany. ACM, 2023. ISBN: 978-1-4503-9422-2. DOI: [10.1145/3544549.3585676](https://doi.org/10.1145/3544549.3585676).
- [173] Colin M. Gray, Cristiana Santos, Nataliia Bielova, Michael Toth, and Damian Clifford. Dark patterns and the legal requirements of consent banners: an interaction criticism perspective. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–18, Yokohama Japan. ACM, May 2021. ISBN: 978-1-4503-8096-6. DOI: [10.1145/3411764.3445779](https://doi.org/10.1145/3411764.3445779).
- [174] Colin M. Gray, Cristiana Teixeira Santos, Nataliia Bielova, and Thomas Mildner. An ontology of dark patterns knowledge: foundations, definitions, and a pathway for shared knowledge-building. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–22, Honolulu HI USA. ACM, May 2024. ISBN: 9798400703300. DOI: [10.1145/3613904.3642436](https://doi.org/10.1145/3613904.3642436).

- [175] Colin M. Gray, Erik Stolterman, and Martin A. Siegel. Reprioritizing the relationship between hci research and practice: bubble-up and trickle-down effects. In *Proceedings of the 2014 conference on Designing interactive systems*, 725–734, Vancouver BC Canada. ACM, June 2014. ISBN: 978-1-4503-2902-6. DOI: [10.1145/2598510.2598595](https://doi.org/10.1145/2598510.2598595).
- [176] Saul Greenberg, Sebastian Boring, Jo Vermeulen, and Jakub Dostal. Dark patterns in proxemic interactions: a critical perspective. *Proceedings of the 2014 Conference on Designing Interactive Systems*. Dis '14:523–532, 2014. DOI: [10.1145/2598510.2598541](https://doi.org/10.1145/2598510.2598541).
- [177] Nancy Greer, Caillem Murray Boyle, and Rebecca Jenkinson. *Harms associated with loot boxes, simulated gambling and other in-game purchases in video games: a review of the evidence*. 2022.
- [178] Anthony W.K. Gaillard G.Robert J. Hockey and Oleksandr Burov, editors. *Operator functional state : the assessment and prediction of human performance : degradation in complex tasks*. English. IOS Press, Netherlands, 2003. ISBN: 4274906086. Pagination: 379.
- [179] Guidelines on transparency under regulation 2016/679. URL: https://www.edpb.europa.eu/our-work-tools/our-documents/article-29-working-party-guidelines-transparency-under-regulation_en.
- [180] Johanna Gunawan, Amogh Pradeep, David Choffnes, Woodrow Hartzog, and Christo Wilson. A comparative study of dark patterns across web and mobile modalities. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW 2):1–29, Oct. 2021. ISSN: 2573-0142. DOI: [10.1145/3479521](https://doi.org/10.1145/3479521).
- [181] Johanna Gunawan, Cristiana Santos, and Irene Kamara. Redress for dark patterns privacy harms? a case study on consent interactions. In *Proceedings of the 2022 Symposium on Computer Science and Law*, 181–194, Washington DC USA. ACM, Nov. 2022. ISBN: 978-1-4503-9234-1. DOI: [10.1145/3511265.3550448](https://doi.org/10.1145/3511265.3550448).
- [182] Meeri Haataja and Joanna J. Bryson. The european parliament's ai regulation: should we call it progress? *Amicus Curiae*, 4(33):707–718, June 2023. ISSN: 2048-481x. DOI: [10.14296/ac.v4i3.5612](https://doi.org/10.14296/ac.v4i3.5612).
- [183] Hana Habib, Megan Li, Ellie Young, and Lorrie Cranor. “okay, whatever”: an evaluation of cookie consent interfaces. In *CHI Conference on Human Factors in Computing Systems*, 1–27, New Orleans LA USA. ACM, Apr. 2022. ISBN: 978-1-4503-9157-3. DOI: [10.1145/3491102.3501985](https://doi.org/10.1145/3491102.3501985).
- [184] Hana Habib, Sarah Pearman, Ellie Young, Ishika Saxena, Robert Zhang, and Lorrie Falth Cranor. Identifying user needs for advertising controls on facebook. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1):1–42, Mar. 2022. ISSN: 2573-0142. DOI: [10.1145/3512906](https://doi.org/10.1145/3512906).
- [185] Hilda Hadan, Lydia Choong, Leah Zhang-Kennedy, and Lennart E. Nacke. Deceived by immersion: a systematic analysis of deceptive design in extended reality. *ACM Computing Surveys*, 56(10):1–25, Oct. 2024. ISSN: 0360-0300, 1557-7341. DOI: [10.1145/3659945](https://doi.org/10.1145/3659945).
- [186] Kim Halskov and Nicolai Brodersen Hansen. The diversity of participatory design research practice at pdc 2002–2012. *International Journal of Human-Computer Studies*, 74:81–92, Feb. 2015. ISSN: 1071-5819. DOI: [10.1016/j.ijhcs.2014.09.003](https://doi.org/10.1016/j.ijhcs.2014.09.003).
- [187] Eszter Hargittai and Amanda Hinnant. Digital inequality: differences in young adults' use of the internet. *Communication Research*, 35(5):602–621, Oct. 2008. DOI: [10.1177/0093650208321782](https://doi.org/10.1177/0093650208321782).
- [188] Julie Harker and Jos Kleijnen. What is a rapid review? a methodological exploration of rapid reviews in health technology assessments. *International Journal of Evidence-Based Healthcare*, 10(4):397–410, Dec. 2012. ISSN: 1744-1595. DOI: [10.1111/j.1744-1609.2012.00290.x](https://doi.org/10.1111/j.1744-1609.2012.00290.x).
- [189] Christina Harrington and Tawanna R Dillahunt. Eliciting tech futures among black young adults: a case study of remote speculative co-design. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–15, Yokohama Japan. ACM, May 2021. ISBN: 978-1-4503-8096-6. DOI: [10.1145/3411764.3445723](https://doi.org/10.1145/3411764.3445723).
- [190] Daniel Harrison, Scarlett Rowland, Gavin Wood, Lyndsey Bakewell, Ioannis Petridis, Kiel Long, Konstantina Vasileiou, Julie Barnett, Manuela Barreto, Michael Wilson, Shaun Lawson, and John Vines. Designing technology-mediated peer support for postgraduate research students at risk of loneliness and isolation. *ACM Transactions on Computer-Human Interaction*, 30(1):1–40, Feb. 2023. ISSN: 1073-0516, 1557-7325. DOI: [10.1145/3534961](https://doi.org/10.1145/3534961).

- [191] Steve Harrison, Deborah Tatar, and Phoebe Sengers. The three paradigms of hci. *Conference: Alt. Chi. Session at the SIGCHI Conference on Human Factors in Computing Systems San Jose, California, USA*, 2007.
- [192] S M Hasan Mansur, Sabiha Salma, Damilola Awofisayo, and Kevin Moran. Aidui: toward automated recognition of dark patterns in user interfaces. In *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*, 1958–1970, Melbourne, Australia. Ieee, May 2023. ISBN: 978-1-66545-701-9. DOI: [10.1109/icse48619.2023.00166](https://doi.org/10.1109/icse48619.2023.00166).
- [193] Janet Heaton. *Reworking Qualitative Data*. SAGE Publications, 2004. ISBN: 9780761971436.
- [194] N. Helberger, M. Sax, J. Strycharz, and H.-W. Micklitz. Choice architectures in the digital economy: towards a new understanding of digital vulnerability. *Journal of Consumer Policy*, 45(2):175–200, 2022. ISSN: 0168-7034, 1573-0700. DOI: [10.1007/s10603-021-09500-5](https://doi.org/10.1007/s10603-021-09500-5).
- [195] Ellen J. Helsper. *The Digital Disconnect. The social causes and consequences of digital inequalities*. SAGE Publications Ltd, 2021.
- [196] Ellen J. Helsper and Alexander J. A. M. van Deursen. Do the rich get digitally richer? quantity and quality of support for digital engagement. *Information, Communication & Society*, 20(5):700–714, May 2017. ISSN: 1369-118x, 1468-4462. DOI: [10.1080/1369118x.2016.1203454](https://doi.org/10.1080/1369118x.2016.1203454).
- [197] Ellen Johanna Helsper. A corresponding fields model for the links between social and digital exclusion: a corresponding fields model for digital exclusion. *Communication Theory*, 22(4):403–426, 2012. ISSN: 10503293. DOI: [10.1111/j.1468-2885.2012.01416.x](https://doi.org/10.1111/j.1468-2885.2012.01416.x).
- [198] Ellen Johanna Helsper. Survey on the use of information and communication technologies in brazilian households: ict households 2015. Núcleo de Informação e Coordenação do Ponto BR:175–185, 2016.
- [199] Ralph Hertwig and Till Grüne-Yanoff. Nudging and boosting: steering or empowering good decisions. *Perspectives on Psychological Science*, 12(6):973–986, Nov. 2017. DOI: [10.1177/1745691617702496](https://doi.org/10.1177/1745691617702496).
- [200] John Hicks and Jeremiah Still. *Examining the Effects of Clutter and Target Salience in an E-Commerce Visual Search Task*, volume 63. Feb. 2020. DOI: [10.1177/1071181319631213](https://doi.org/10.1177/1071181319631213). Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
- [201] Shun Hidaka, Sota Kobuki, Mizuki Watanabe, and Katie Seaborn. 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*, 1–13, Hamburg Germany. ACM, Apr. 2023. ISBN: 978-1-4503-9421-5. DOI: [10.1145/3544548.3580942](https://doi.org/10.1145/3544548.3580942).
- [202] Gerard P. Hodgkinson, Nicola J. Bown, A. John Maule, Keith W. Glaister, and Alan D. Pearman. Breaking the frame: an analysis of strategic cognition and decision making under uncertainty. *Strategic Management Journal*, 20(10):977–985, 1999. ISSN: 0143-2095.
- [203] Mairéad Hogan, Chris Barry, and Michael Lang. Dissecting optional micro-decisions in online transactions: perceptions, deceptions and errors. *ACM Transactions on Computer-Human Interaction*:3531005, May 2022. ISSN: 1073-0516, 1557-7325. DOI: [10.1145/3531005](https://doi.org/10.1145/3531005).
- [204] Merlyn Holkar and Chris Lees. *A Safer Bet? Online Gambling and Mental Health*. 2020. URL: https://www.moneyandmentalhealth.org/wp-content/uploads/2020/07/A_Safer_Bet.pdf.pdf.
- [205] Linda D. Hollebeek, V. Kumar, Rajendra K. Srivastava, and Moira K. Clark. Moving the stakeholder journey forward. *Journal of the Academy of Marketing Science*, 51(1):23–49, Jan. 2023. ISSN: 1552-7824. DOI: [10.1007/s11747-022-00878-3](https://doi.org/10.1007/s11747-022-00878-3).
- [206] Kat Holmes. *The cycle of exclusion*. In *Mismatch. How Inclusion Shapes Design*. MIT Press, 2020, 27–40.
- [207] Kasper Hornbæk and Antti Oulasvirta. What is interaction? In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Chi '17, 5040–5052, New York, NY, USA. Association for Computing Machinery, May 2017. ISBN: 978-1-4503-4655-9. DOI: [10.1145/3025453.3025765](https://doi.org/10.1145/3025453.3025765).
- [208] Hsiu-Fang Hsieh and Sarah E Shannon. Three approaches to qualitative content analysis. *Qualitative health research*, 15(9):1277–1288, Nov. 2005. ISSN: 1049-7323. DOI: [10.1177/1049732305276687](https://doi.org/10.1177/1049732305276687).
- [209] Nobuaki Hsu and Zacharias Schütt. *Psychology of priming*. Aug. 2012, page 272.
- [210] Kahryn Hughes and Anna Tarrant. *Qualitative Secondary Analysis*. SAGE Publications, 2019. ISBN: 9781526482860.

- [211] Lujain Ibrahim, Luc Rocher, and Ana Valdivia. Characterizing and modeling harms from interactions with design patterns in ai interfaces. (arXiv:2404.11370), May 2024. DOI: [10.48550/arXiv.2404.11370](https://doi.org/10.48550/arXiv.2404.11370). arXiv:2404.11370 [cs].
- [212] Grace Lorraine Intal, Gabe Andre Llena, Patrick Leus Vidal, and Angelica Young. System analysis and design of an automated charity donation system using design thinking methodology. In *Proceedings of the 2023 8th International Conference on Information Systems Engineering*, Icise '23, 29–35, New York, NY, USA. Association for Computing Machinery, June 2024. ISBN: 9798400709173. DOI: [10.1145/3641032.3641048](https://doi.org/10.1145/3641032.3641048).
- [213] Jennifer Ismirle. Using experience maps to consider individual stories. In *Proceedings of the 36th ACM International Conference on the Design of Communication*, Sigdoc '18, 1–6, New York, NY, USA. Association for Computing Machinery, Aug. 2018. ISBN: 978-1-4503-5935-1. DOI: [10.1145/3233756.3233954](https://doi.org/10.1145/3233756.3233954).
- [214] Naomi Jacobs. Capability sensitive design for health and wellbeing technologies. *Science and Engineering Ethics*, 26(6):3363–3391, Dec. 2020. ISSN: 1353-3452, 1471-5546. DOI: [10.1007/s11948-020-00275-5](https://doi.org/10.1007/s11948-020-00275-5).
- [215] Elaheh Jafari and Julita Vassileva. Designing effective warnings for manipulative designs in mobile applications. In *Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization*, 12–17, Cagliari Italy. ACM, June 2024. ISBN: 9798400704338. DOI: [10.1145/3627043.3659550](https://doi.org/10.1145/3627043.3659550).
- [216] Luiza Jarovsky. Dark patterns in personal data collection: definition, taxonomy and lawfulness. (4048582), Mar. 2022. DOI: [10.2139/ssrn.4048582](https://doi.org/10.2139/ssrn.4048582).
- [217] Sirkka L. Jarvenpaa. Graphic displays in decision making — the visual salience effect. *Journal of Behavioral Decision Making*, 3(4):247–262, 1990. ISSN: 1099-0771. DOI: [10.1002/bdm.3960030403](https://doi.org/10.1002/bdm.3960030403).
- [218] Haiyan Jia, Pamela J. Wisniewski, Heng Xu, Mary Beth Rosson, and John M. Carroll. Risk-taking as a learning process for shaping teen’s online information privacy behaviors. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 583–599, Vancouver BC Canada. ACM, 2015. ISBN: 978-1-4503-2922-4. DOI: [10.1145/2675133.2675287](https://doi.org/10.1145/2675133.2675287).
- [219] Xiong Jiang, Yang Jiang, and Raja Parasuraman. The visual priming of motion-defined 3d objects. *Plos One*, 10(12):e0144730, Dec. 2015. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0144730](https://doi.org/10.1371/journal.pone.0144730).
- [220] Genevieve Marie Johnson and Korbla P. Pupilampu. Internet use during childhood and the ecological techno-subsystem. *Canadian Journal of Learning and Technology / La revue canadienne de l'apprentissage et de la technologie*, 34(1), Dec. 2008. ISSN: 1499-6685, 1499-6677. DOI: [10.21432/t2cp4t](https://doi.org/10.21432/t2cp4t).
- [221] A N Joinson, D J Houghton, A Vassalou, and B L Marder. Digital crowding:privacy, self-disclosure:8.
- [222] Meg Leta Jones. Cookies: a legacy of controversy. *Internet Histories*, 4(1):87–104, Jan. 2020. ISSN: 2470-1475, 2470-1483. DOI: [10.1080/24701475.2020.1725852](https://doi.org/10.1080/24701475.2020.1725852).
- [223] Daniel Kahneman. *Thinking, fast and slow*. English. Penguin, London, 2012. ISBN: 9780141033570 0141033576.
- [224] Kanav Kahol, Mario J. Leyba, Mary Deka, Vikram Deka, Stephanie Mayes, Marshall Smith, John J. Ferrara, and Sethuraman Panchanathan. Effect of fatigue on psychomotor and cognitive skills. eng. *American Journal of Surgery*, 195(2):195–204, Feb. 2008. ISSN: 1879-1883. DOI: [10.1016/j.amjsurg.2007.10.004](https://doi.org/10.1016/j.amjsurg.2007.10.004).
- [225] Jim Kalbach. *Mapping Experiences: A Complete Guide to Creating Value through Journeys, Blueprints, and Diagrams*. O'Reilly Media, 2016. ISBN: 9781491923481.
- [226] Georgios Kampanos and Siamak Shahandashti. *Accept AllFIX ME!!!!: The Landscape of Cookie Banners in Greece and the UK*. Mar. 2021.
- [227] Stanley Kaplan and B. John Garrick. On the quantitative definition of risk. *Risk Analysis*, 1(1):11–27, 1981. ISSN: 1539-6924. DOI: [10.1111/j.1539-6924.1981.tb01350.x](https://doi.org/10.1111/j.1539-6924.1981.tb01350.x).
- [228] Ilayda Karagoel and Dan Nathan-Roberts. Dark patterns: social media, gaming, and e-commerce. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 65(1):752–756, Sept. 2021. ISSN: 2169-5067, 1071-1813. DOI: [10.1177/1071181321651317](https://doi.org/10.1177/1071181321651317).
- [229] Ioannis Kareklas, Darrel D. Muehling, and Skyler King. The effect of color and self-view priming in persuasive communications. *Journal of Business Research*, 98:33–49, May 2019. ISSN: 0148-2963. DOI: [10.1016/j.jbusres.2019.01.022](https://doi.org/10.1016/j.jbusres.2019.01.022).

- [230] Maxwell Keleher, Fiona Westin, Preethi Nagabandi, and Sonia Chiasson. How well do experts understand end-users' perceptions of manipulative patterns? In *Nordic Human-Computer Interaction Conference*, NordiCHI '22, 1–21, Aarhus Denmark. ACM, Oct. 2022. ISBN: 978-1-4503-9699-8. DOI: [10.1145/3546155.3546656](https://doi.org/10.1145/3546155.3546656).
- [231] Dominique Kelly and Victoria L Rubin. Dark pattern typology: how do social networking sites deter disabling of user accounts?:5, 2022. DOI: [10.13140/rg.2.2.14087.47528](https://doi.org/10.13140/rg.2.2.14087.47528).
- [232] Kay Kender and Christopher Frauenberger. The shape of social media: towards addressing (aesthetic) design power. In *Designing Interactive Systems Conference*, 365–376, Virtual Event Australia. ACM, June 2022. ISBN: 978-1-4503-9358-4. DOI: [10.1145/3532106.3533470](https://doi.org/10.1145/3532106.3533470).
- [233] Maximilian Kiener. When do nudges undermine voluntary consent? *Philosophical Studies*, 178(12):4201–4226, Dec. 2021. ISSN: 0031-8116, 1573-0883. DOI: [10.1007/s11098-021-01644-x](https://doi.org/10.1007/s11098-021-01644-x).
- [234] Woo Gon Kim, Souji Gopalakrishna Pillai, Kavitha Haldorai, and Wasim Ahmad. Dark patterns used by online travel agency websites. *Annals of Tourism Research*, 88:103055, May 2021. ISSN: 01607383. DOI: [10.1016/j.annals.2020.103055](https://doi.org/10.1016/j.annals.2020.103055).
- [235] Daniel Kirkman, Kami Vaniea, and Daniel W. Woods. Darkdialogs: automated detection of 10 dark patterns on cookie dialogs. English. In 847–867. IEEE Computer Society, July 2023. ISBN: 978-1-66546-512-0. DOI: [10.1109/EuroSP57164.2023.00055](https://doi.org/10.1109/EuroSP57164.2023.00055).
- [236] Agnieszka Kitkowska, Johan Högberg, and Erik Wästlund. Barriers to a well-functioning digital market: exploring dark patterns and how to overcome them. In 2022. DOI: [10.24251/hicss.2022.573](https://doi.org/10.24251/hicss.2022.573).
- [237] Jennifer Klütsch, Christian Böffel, Sophia Von Salm-Hoogstraeten, and Sabine J. Schlittmeier. *Defeating dark patterns: the impact of supporting information on dark patterns and cookie privacy decisions*. In *International Symposium on Technikpsychologie (TecPsy) 2023*. Nina Gerber and Verena Zimmermann, editors. Sciendo, June 2023, 41–54. ISBN: 978-83-66675-89-6. DOI: [10.2478/9788366675896-004](https://doi.org/10.2478/9788366675896-004).
- [238] Ellen Kohl and Priscilla McCutcheon. Kitchen table reflexivity: negotiating positionality through everyday talk. *Gender, Place & Culture*, 22(6):747–763, July 2015. ISSN: 0966-369x, 1360-0524. DOI: [10.1080/0966369x.2014.958063](https://doi.org/10.1080/0966369x.2014.958063).
- [239] Tim Kollmer. Exploring the impact of digital sludging. In *PACIS 2022 Proceedings*, page 10, 2022.
- [240] Tim Kollmer and Andreas Eckhardt. Dark patterns: conceptualization and future research directions. *Business & Information Systems Engineering*, 65(2):201–208, Apr. 2023. ISSN: 2363-7005, 1867-0202. DOI: [10.1007/s12599-022-00783-7](https://doi.org/10.1007/s12599-022-00783-7).
- [241] Konrad Kollnig, Siddhartha Datta, and Max Van Kleek. I want my app that way: reclaiming sovereignty over personal devices. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–8, Yokohama Japan. ACM, May 2021. ISBN: 978-1-4503-8095-9. DOI: [10.1145/3411763.3451632](https://doi.org/10.1145/3411763.3451632).
- [242] Irene Korstjens and Albine Moser. Series: practical guidance to qualitative research. part 4: trustworthiness and publishing. en. *European Journal of General Practice*, 24(1):120–124, Jan. 2018. ISSN: 1381-4788, 1751-1402. DOI: [10.1080/13814788.2017.1375092](https://doi.org/10.1080/13814788.2017.1375092).
- [243] Monica Kowalczyk, Johanna T. Gunawan, David Choffnes, Daniel J Dubois, Woodrow Hartzog, and Christo Wilson. Understanding dark patterns in home iot devices. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Chi '23, 1–27, Hamburg Germany. ACM, Apr. 2023. ISBN: 978-1-4503-9421-5. DOI: [10.1145/3544548.3581432](https://doi.org/10.1145/3544548.3581432).
- [244] Anastasia Kozyreva, Stephan Lewandowsky, and Ralph Hertwig. Citizens versus the internet: confronting digital challenges with cognitive tools. eng. *Psychological Science in the Public Interest: A Journal of the American Psychological Society*, 21(3):103–156, Dec. 2020. ISSN: 2160-0031. DOI: [10.1177/1529100620946707](https://doi.org/10.1177/1529100620946707).
- [245] Veronika Krauß. Exploring dark patterns in xr:2, 2022. URL: https://www.researchgate.net/publication/359578610_Exploring_Dark_Patterns_in_XR.
- [246] Veronika Krauß, Pejman Saeghe, Alexander Boden, Mohamed Khamis, Mark McGill, Jan Gugenheimer, and Michael Nebeling. What makes xr dark? examining emerging dark patterns in augmented and virtual reality through expert co-design. *ACM Transactions on Computer-Human Interaction*, 31(3):1–39, June 2024. ISSN: 1073-0516, 1557-7325. DOI: [10.1145/3660340](https://doi.org/10.1145/3660340).

- [247] Chiara Krisam, Heike Dietmann, Melanie Volkamer, and Oksana Kulyk. Dark patterns in the wild: review of cookie disclaimer designs on top 500 german websites. In *European Symposium on Usable Security 2021*, 1–8, Karlsruhe Germany. ACM, Oct. 2021. ISBN: 978-1-4503-8423-0. DOI: [10.1145/3481357.3481516](https://doi.org/10.1145/3481357.3481516).
- [248] Max Krüger, Anne Weibert, Debora De Castro Leal, Dave Randall, and Volker Wulf. “what is the topic of the group, please?” on migration, care and the challenges of participation in design. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–29, Oct. 2021. ISSN: 2573-0142. DOI: [10.1145/3476050](https://doi.org/10.1145/3476050).
- [249] Max Krüger, Anne Weibert, Debora De Castro Leal, Dave Randall, and Volker Wulf. It takes more than one hand to clap: on the role of ‘care’ in maintaining design results. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–14, Yokohama Japan. ACM, May 2021. ISBN: 978-1-4503-8096-6. DOI: [10.1145/3411764.3445389](https://doi.org/10.1145/3411764.3445389). URL: <https://dl.acm.org/doi/10.1145/3411764.3445389>.
- [250] Oksana Kulyk, Nina Gerber, Annika Hilt, and Melanie Volkamer. Has the gdpr hype affected users’ reaction to cookie disclaimers? *Journal of Cybersecurity*, 6(1):tyaa022, Jan. 2020. ISSN: 2057-2085. DOI: [10.1093/cybsec/tyaa022](https://doi.org/10.1093/cybsec/tyaa022).
- [251] Oksana Kulyk, Annika Hilt, Nina Gerber, and Melanie Volkamer. “This Website Uses Cookies”: Users’ Perceptions and Reactions to the Cookie Disclaimer. Apr. 2018. DOI: [10.14722/eurousec.2018.23012](https://doi.org/10.14722/eurousec.2018.23012).
- [252] Lin Kyi, Sushil Ammanaghatta Shivakumar, Cristiana Teixeira Santos, Franziska Roesner, Frederike Zufall, and Asia J. Biega. Investigating deceptive design in gdpr’s legitimate interest. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Chi ’23, 1–16, New York, NY, USA. Association for Computing Machinery, Apr. 2023. ISBN: 978-1-4503-9421-5. DOI: [10.1145/3544548.3580637](https://doi.org/10.1145/3544548.3580637).
- [253] Stephan Käufer and Anthony Chemero. *James J. Gibson and ecological psychology*. In *Phenomenology: An introduction*. Wiley, 2021, 181–199. ISBN: 978-1-5095-4066-2.
- [254] Cherie Lacey, Alex Beattie, and Tristram Sparks. Clusters of dark patterns across popular websites in new zealand. *International Journal of Communication*, 17(00):25, Apr. 2023. ISSN: 1932-8036.
- [255] Cherie Lacey and Catherine Caudwell. Cuteness as a ‘dark pattern’ in home robots. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 374–381, Daegu, Korea (South). Ieee, Mar. 2019. ISBN: 978-1-5386-8555-6. DOI: [10.1109/hri.2019.8673274](https://doi.org/10.1109/hri.2019.8673274).
- [256] Carine Lallemand, Guillaume Gronier, and Marc Dugué. *Méthodes de design UX: 30 méthodes fondamentales pour concevoir des expériences optimales*. fr. Design Web. Eyrolles, Paris, 2e éd edition, 2018. ISBN: 978-2-212-67398-2.
- [257] Carine Lallemand and Vincent Koenig. Measuring the contextual dimension of user experience: development of the user experience context scale (uxcs). In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, 1–13, Tallinn Estonia. ACM, Oct. 2020. ISBN: 978-1-4503-7579-5. DOI: [10.1145/3419249.3420156](https://doi.org/10.1145/3419249.3420156).
- [258] Carine Lallemand, Jessie Lauret, and Luce Drouet. Physical journey maps: staging users’ experiences to increase stakeholders’ empathy towards users. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, Chi Ea ’22, 1–7, New York, NY, USA. Association for Computing Machinery, Apr. 2022. ISBN: 978-1-4503-9156-6. DOI: [10.1145/3491101.3519630](https://doi.org/10.1145/3491101.3519630).
- [259] Tony C M Lam and Priscilla Bengo. A comparison of three retrospective self-reporting methods of measuring change in instructional practice. *American Journal Of Evaluation*, 2003.
- [260] Marjolein Lanzing. ‘strongly recommended’ revisiting decisional privacy to judge hypernudging in self-tracking technologies. *Philos. Technol.*, (32):549–568, 2019.
- [261] Kristina Lapin and Laima Volungevičiūtė. *Improving the usability of requests for consent to use cookies*. In *Digital Interaction and Machine Intelligence*. Cezary Biele, Janusz Kacprzyk, Wiesław Kopeć, Jan W. Owsński, Andrzej Romanowski, and Marcin Sikorski, editors. Volume 710. Lecture Notes in Networks and Systems. Springer Nature Switzerland, Cham, 2023, 191–201. ISBN: 978-3-031-37648-1. DOI: [10.1007/978-3-031-37649-8_19](https://doi.org/10.1007/978-3-031-37649-8_19).
- [262] Celine Latulipe, Ronnie Dsouza, and Murray Cumbers. Unofficial proxies: how close others help older adults with banking. In *CHI Conference on Human Factors in Computing Systems*, 1–13, New Orleans LA USA. ACM, Apr. 2022. ISBN: 978-1-4503-9157-3. DOI: [10.1145/3491102.3501845](https://doi.org/10.1145/3491102.3501845).

- [263] Johann Laux, Sandra Wachter, and Brent Mittelstadt. Trustworthy artificial intelligence and the european union ai act: on the conflation of trustworthiness and acceptability of risk. *Regulation & Governance*, 18(1):3–32, 2024. ISSN: 1748-5991. DOI: [10.1111/rego.12512](https://doi.org/10.1111/rego.12512).
- [264] Susann M. Lavery. Hermeneutic phenomenology and phenomenology: a comparison of historical and methodological considerations. *International Journal of Qualitative Methods*, 2(3):21–35, Sept. 2003. ISSN: 1609-4069, 1609-4069. DOI: [10.1177/160940690300200303](https://doi.org/10.1177/160940690300200303).
- [265] Cheol-Soon Lee and Kwame McKenzie. Socioeconomic and geographic inequalities of internet addiction in korean adolescents. *Psychiatry Investigation*, 12(4):559–562, Oct. 2015. ISSN: 1738-3684. DOI: [10.4306/pi.2015.12.4.559](https://doi.org/10.4306/pi.2015.12.4.559).
- [266] Hyunwoo Lee, Ayoung Cho, Jaewon Jung, Rayoung Park, and Mincheol Whang. Customer journey map system based on lifelogging. In *Proceedings of HCI Korea, Hcik '16*, 266–272, Seoul, KOR. Hanbit Media, Inc., Jan. 2016. ISBN: 978-89-6848-791-0. DOI: [10.17210/hcik.2016.01.266](https://doi.org/10.17210/hcik.2016.01.266).
- [267] M. R. Leiser and Cristiana Santos. Dark patterns, enforcement, and the emerging digital design acquis: manipulation beneath the interface. (4431048), Apr. 2023. URL: <https://papers.ssrn.com/abstract=4431048>.
- [268] Giacomo Lepri and Andrew McPherson. Making up instruments: design fiction for value discovery in communities of musical practice. In *Proceedings of the 2019 on Designing Interactive Systems Conference*, 113–126, San Diego CA USA. ACM, June 2019. ISBN: 978-1-4503-5850-7. DOI: [10.1145/3322276.3322353](https://doi.org/10.1145/3322276.3322353).
- [269] Lawrence Lessig. *Code and Other Laws of Cyberspace*. Basic Books, Inc., Usa, Sept. 1999. ISBN: 978-0-465-03912-8.
- [270] Danyang Li. The ftc and the cpra's regulation of dark patterns in cookie consent notices. *The University of Chicago Business Law Review*, 1.
- [271] Lingyuan Li, Guo Freeman, and Donghee Yvette Wohn. Power in skin: the interplay of self-presentation, tactical play, and spending in fortnite. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 71–80, Virtual Event Canada. ACM, Nov. 2020. ISBN: 978-1-4503-8074-4. DOI: [10.1145/3410404.3414262](https://doi.org/10.1145/3410404.3414262).
- [272] Pranee Liamputtong. Researching the vulnerable: a guide to sensitive research methods:1–256, 2006. DOI: <https://doi.org/10.4135/9781849209861>.
- [273] Sonia Livingstone and Peter K. Smith. Annual research review: harms experienced by child users of online and mobile technologies: the nature, prevalence and management of sexual and aggressive risks in the digital age. *Journal of Child Psychology and Psychiatry*, 55(6):635–654, 2014. ISSN: 00219630. DOI: [10.1111/jcpp.12197](https://doi.org/10.1111/jcpp.12197).
- [274] Sonia Livingstone, Mariya Stoilova, and Rishita Nandagiri. *Children's data and privacy online: Growing up in a digital age*. 2019, 153–173. URL: <https://scindeks.ceon.rs/Article.aspx?artid=2334-959X2301153D>.
- [275] Dan Lockton. Persuasive technology and digital design for behaviour change. (2125957), Aug. 2012. DOI: [10.2139/ssrn.2125957](https://doi.org/10.2139/ssrn.2125957).
- [276] Dan Lockton, David Harrison, and Neville Stanton. *Design with intent: persuasive technology in a wider context*. In *Persuasive Technology*. Harri Oinas-Kukkonen, Per Hasle, Marja Harjumaa, Katarina Segerståhl, and Peter Øhrstrøm, editors. Volume 5033. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, 274–278. ISBN: 978-3-540-68500-5. DOI: [10.1007/978-3-540-68504-3_30](https://doi.org/10.1007/978-3-540-68504-3_30).
- [277] Kiel Long, Lyndsey L. Bakewell, Roisin C. McNaney, Konstantina Vasileiou, Mark Atkinson, Manuela Barreto, Julie Barnett, Michael Wilson, Shaun Lawson, and John Vines. Connecting those that care: designing for transitioning, talking, belonging and escaping. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 1339–1351, Denver Colorado USA. ACM, May 2017. ISBN: 978-1-4503-4655-9. DOI: [10.1145/3025453.3025715](https://doi.org/10.1145/3025453.3025715).
- [278] Mengyi Long, Yue Xu, Jiangrong Wu, Qihua Ou, and Yuhong Nan. Understanding dark ui patterns in the mobile ecosystem: a case study of apps in china. In *Proceedings of the 2023 ACM Workshop on Secure and Trustworthy Superapps*, 33–40, Copenhagen Denmark. ACM, Nov. 2023. ISBN: 9798400702587. DOI: [10.1145/3605762.3624431](https://doi.org/10.1145/3605762.3624431).

- [279] Yuwen Lu, Chao Zhang, Yuewen Yang, Yaxing Yao, and Toby Jia-Jun Li. From awareness to action: exploring end-user empowerment interventions for dark patterns in ux. (arXiv:2310.17846), Feb. 2024. URL: <http://arxiv.org/abs/2310.17846>. arXiv:2310.17846 [cs].
- [280] Jamie Luguri and Lior Jacob Strahilevitz. Shining a light on dark patterns. *Journal of Legal Analysis*, 13(1):43–109, Mar. 2021. ISSN: 2161-7201, 1946-5319. DOI: [10.1093/jla/laaa006](https://doi.org/10.1093/jla/laaa006).
- [281] Kai Lukoff, Ulrik Lyngs, Himanshu Zade, J. Vera Liao, James Choi, Kaiyue Fan, Sean A. Munson, and Alexis Hiniker. How the design of youtube influences user sense of agency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–17, Yokohama Japan. ACM, May 2021. ISBN: 978-1-4503-8096-6. DOI: [10.1145/3411764.3445467](https://doi.org/10.1145/3411764.3445467).
- [282] Kai Lukoff, Ulrik Lyngs, Himanshu Zade, J Vera Liao, James Choi, Kaiyue Fan, Sean A Munson, and Alexis Hiniker. How the design of YouTube influences user sense of agency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, number Article 368 in Chi '21, pages 1–17, Yokohama, Japan. Association for Computing Machinery, May 2021. ISBN: 9781450380966. DOI: [10.1145/3411764.3445467](https://doi.org/10.1145/3411764.3445467).
- [283] Florencia Luna. Identifying and evaluating layers of vulnerability – a way forward. *Developing World Bioethics*, 19(2):86–95, June 2019. ISSN: 1471-8731, 1471-8847. DOI: [10.1111/dewb.12206](https://doi.org/10.1111/dewb.12206).
- [284] Francisco Lupiáñez Villanueva, Alba Boluda, Francesco Bogliacino, and Giovanni Liva. Behavioural study on unfair commercial practices in the digital environment: dark patterns and manipulative personalisation:303, 2022.
- [285] Ulrik Lyngs, Kai Lukoff, Petr Slovak, William Seymour, Helena Webb, Marina Jirotko, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. “i just want to hack myself to not get distracted”: evaluating design interventions for self-control on facebook. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15, Apr. 2020. DOI: [10.1145/3313831.3376672](https://doi.org/10.1145/3313831.3376672). arXiv:2001.04180 [cs].
- [286] Aditi M. Bhoot, Mayuri A. Shinde, and Wricha P. Mishra. Towards the identification of dark patterns: an analysis based on end-user reactions. In *IndiaHCI '20: Proceedings of the 11th Indian Conference on Human-Computer Interaction*, 24–33, Online India. ACM, Nov. 2020. ISBN: 978-1-4503-8944-0. DOI: [10.1145/3429290.3429293](https://doi.org/10.1145/3429290.3429293).
- [287] Maximilian Maier and Rikard Harr. Dark design patterns: an end-user perspective. *Human Technology*, 16(2):170–199, Aug. 2020. ISSN: 17956889. DOI: [10.17011/ht/urn.202008245641](https://doi.org/10.17011/ht/urn.202008245641).
- [288] Gianclaudio Malgieri. The vulnerable data subject in the gdpr. In *Vulnerability and Data Protection Law*, 1st edition. Oxford University Press, Apr. 2023. DOI: [10.1093/oso/9780192870339.003.0004](https://doi.org/10.1093/oso/9780192870339.003.0004).
- [289] Gianclaudio Malgieri. Who is the vulnerable individual? In *Vulnerability and Data Protection Law*, C3–c3n158, 1st edition. Oxford University Press, Apr. 2023. DOI: [10.1093/oso/9780192870339.003.0003](https://doi.org/10.1093/oso/9780192870339.003.0003).
- [290] Gianclaudio Malgieri and Jkedorzrej Niklas. Vulnerable data subjects. *Computer Law & Security Review*, 37:105415, 2020. ISSN: 02673649. DOI: [10.1016/j.clsr.2020.105415](https://doi.org/10.1016/j.clsr.2020.105415).
- [291] Alessandro Mantelero. Ai and big data: a blueprint for a human rights, social and ethical impact assessment. *Computer Law & Security Review*, 34(4):754–772, Aug. 2018. ISSN: 02673649. DOI: [10.1016/j.clsr.2018.05.017](https://doi.org/10.1016/j.clsr.2018.05.017).
- [292] Alessandro Mantelero. Fundamental rights impact assessments in the dsa. eng. *Verfassungsblog*, Nov. 2022. ISSN: 2366-7044. DOI: [10.17176/20221101-220006-0](https://doi.org/10.17176/20221101-220006-0). URL: <https://verfassungsblog.de/dsa-impact-assessment/>.
- [293] Alessandro Mantelero. The fundamental rights impact assessment (fria) in the ai act: roots, legal obligations and key elements for a model template. (4782126), Mar. 2024. DOI: [10.2139/ssrn.4782126](https://doi.org/10.2139/ssrn.4782126).
- [294] Danica Mast, Joost Broekens, Sanne I. de Vries, and Fons J. Verbeek. Participation patterns of interactive playful museum exhibits: evaluating the participant journey map through situated observations. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, Dis '23, 1861–1885, New York, NY, USA. Association for Computing Machinery, July 2023. ISBN: 978-1-4503-9893-0. DOI: [10.1145/3563657.3595985](https://doi.org/10.1145/3563657.3595985).
- [295] Arunesh Mathur. Identifying and measuring manipulative user interfaces at scale on the web. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–5, Yokohama Japan. ACM, May 2021. ISBN: 978-1-4503-8095-9. DOI: [10.1145/3411763.3457782](https://doi.org/10.1145/3411763.3457782).

- [296] Arunesh Mathur, Gunes Acar, Michael J. Friedman, Eli Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. Dark patterns at scale: findings from a crawl of 11k shopping websites. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–32, Nov. 2019. ISSN: 2573-0142. DOI: [10.1145/3359183](https://doi.org/10.1145/3359183). arXiv: 1907.07032.
- [297] Arunesh Mathur, Mihir Kshirsagar, and Jonathan Mayer. What makes a dark pattern... dark?: design attributes, normative considerations, and measurement methods. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–18, Yokohama Japan. ACM, May 2021. ISBN: 978-1-4503-8096-6. DOI: [10.1145/3411764.3445610](https://doi.org/10.1145/3411764.3445610). arXiv: 2101.04843.
- [298] Célestin Matte, Nataliia Bielova, and Cristiana Santos. Do cookie banners respect my choice? measuring legal compliance of banners from iab europe's transparency and consent framework. *arXiv:1911.09964 [cs]*, Feb. 2020. URL: <http://arxiv.org/abs/1911.09964>. arXiv: 1911.09964.
- [299] Gerald Matthews, P.A. Desmond, Catherine Neubauer, and Peter Hancock. *The Handbook of Operator Fatigue*. Jan. 2012, page 510. ISBN: 978-1-315-55736-6. DOI: [10.1201/9781315557366](https://doi.org/10.1201/9781315557366).
- [300] Nora McDonald and Andrea Forte. The politics of privacy theories: moving from norms to vulnerabilities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14, Honolulu HI USA. ACM, Apr. 2020. ISBN: 978-1-4503-6708-0. DOI: [10.1145/3313831.3376167](https://doi.org/10.1145/3313831.3376167).
- [301] Elisa Menardo, Giulia Balboni, and Roberto Cubelli. Environmental factors and teenagers' personalities: the role of personal and familial socio-cultural level. *Behavioural Brain Research*, 325:181–187, 2017. ISSN: 01664328. DOI: [10.1016/j.bbr.2017.02.038](https://doi.org/10.1016/j.bbr.2017.02.038).
- [302] Christopher Mendez, Lara Letaw, Margaret Burnett, Simone Stumpf, Anita Sarma, and Claudia Hilderbrand. From gendermag to inclusivemag: an inclusive design meta-method. In *2019 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 97–106, Oct. 2019. DOI: [10.1109/vlhcc.2019.8818889](https://doi.org/10.1109/vlhcc.2019.8818889).
- [303] Daphne Menheere, Carine Lallemand, Erik Van Der Spek, Carl Megens, Andrew Vande Moere, Mathias Funk, and Steven Vos. The runner's journey: identifying design opportunities for running motivation technology. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, 1–14, Tallinn Estonia. ACM, Oct. 2020. ISBN: 978-1-4503-7579-5. DOI: [10.1145/3419249.3420151](https://doi.org/10.1145/3419249.3420151).
- [304] Christian Meske and Ireti Amojó. Ethical guidelines for the construction of digital nudges. In *Proceedings on the 53rd Hawaii International Conference on System Sciences*, 3928–3937, 2020.
- [305] Marisa Meyer, Victoria Adkins, Nalingna Yuan, Heidi M. Weeks, Yung-Ju Chang, and Jenny Radesky. Advertising in young children's apps: a content analysis. *Journal of developmental and behavioral pediatrics: JDBP*, 40(1):32–39, Jan. 2019. ISSN: 1536-7312. DOI: [10.1097/dbp.0000000000000622](https://doi.org/10.1097/dbp.0000000000000622).
- [306] 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. Listening to the voices: describing ethical caveats of conversational user interfaces according to experts and frequent users. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–18, Honolulu HI USA. ACM, May 2024. ISBN: 9798400703300. DOI: [10.1145/3613904.3642542](https://doi.org/10.1145/3613904.3642542).
- [307] Thomas Mildner, Philip Doyle, Gian-Luca Savino, and Rainer Malaka. Rules of engagement: levelling up to combat unethical cui design. In *Proceedings of the 4th Conference on Conversational User Interfaces*, 1–5, Glasgow United Kingdom. ACM, July 2022. ISBN: 978-1-4503-9739-1. DOI: [10.1145/3543829.3544528](https://doi.org/10.1145/3543829.3544528).
- [308] Thomas Mildner, Merle Freye, Gian-Luca Savino, Philip R. Doyle, Benjamin R. Cowan, and Rainer Malaka. Defending against the dark arts: recognising dark patterns in social media. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, 2362–2374, Pittsburgh PA USA. ACM, July 2023. ISBN: 978-1-4503-9893-0. DOI: [10.1145/3563657.3595964](https://doi.org/10.1145/3563657.3595964).
- [309] Thomas Mildner and Gian-Luca Savino. Ethical user interfaces: exploring the effects of dark patterns on facebook. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–7, Yokohama Japan. ACM, May 2021. ISBN: 978-1-4503-8095-9. DOI: [10.1145/3411763.3451659](https://doi.org/10.1145/3411763.3451659).
- [310] Thomas Mildner, Gian-Luca Savino, Philip R. Doyle, Benjamin R. Cowan, and Rainer Malaka. About engaging and governing strategies: a thematic analysis of dark patterns in social networking services. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–15, Hamburg Germany. ACM, Apr. 2023. ISBN: 978-1-4503-9421-5. DOI: [10.1145/3544548.3580695](https://doi.org/10.1145/3544548.3580695).

- [311] Milica Milosavljevic, Vidhya Navalpakkam, Christof Koch, and Antonio Rangel. Relative visual saliency differences induce sizable bias in consumer choice. *Journal of Consumer Psychology*, 22(1):67–74, 2012. ISSN: 1057-7408.
- [312] Alexander Mirnig and Manfred Tscheligi. (don't) join the dark side: an initial analysis and classification of regular, anti-, and dark patterns. English. In *PATTERNS 2017: Proceedings of the 9th International Conference on Pervasive Patterns and Applications*, pages 65–71. IARIA XPS Press, 2017.
- [313] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G Altman, and PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*, 151(4):264–9, W64, Aug. 2009. ISSN: 0003-4819, 1539-3704. DOI: [10 . 7326 / 0003 - 4819 - 151 - 4 - 200908180 - 00135](https://doi.org/10.7326/0003-4819-151-4-200908180-00135).
- [314] Alberto Monge Roffarello and Luigi De Russis. Towards understanding the dark patterns that steal our attention. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 1–7, New Orleans LA USA. ACM, Apr. 2022. ISBN: 978-1-4503-9156-6. DOI: [10 . 1145/3491101 . 3519829](https://doi.org/10.1145/3491101.3519829).
- [315] Alberto Monge Roffarello, Kai Lukoff, and Luigi De Russis. Defining and identifying attention capture deceptive designs in digital interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–19, Hamburg Germany. ACM, Apr. 2023. ISBN: 978-1-4503-9421-5. DOI: [10 . 1145 / 3544548 . 3580729](https://doi.org/10.1145/3544548.3580729).
- [316] Carol Moser, Sarita Y. Schoenebeck, and Paul Resnick. Impulse buying: design practices and consumer needs. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–15, Glasgow Scotland Uk. ACM, May 2019. ISBN: 978-1-4503-5970-2. DOI: [10 . 1145/3290605 . 3300472](https://doi.org/10.1145/3290605.3300472).
- [317] Jerad Moxley, Joseph Sharit, and Sara J Czaja. The factors influencing older adults' decisions surrounding adoption of technology: quantitative experimental study. *JMIR Aging*, 5(4):e39890, Nov. 2022. ISSN: 2561-7605. DOI: [10 . 2196/39890](https://doi.org/10.2196/39890).
- [318] Diego Muñoz, Bernd Ploderer, and Margot Brereton. Position exchange workshops: a method to design for each other in families. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–14, Glasgow Scotland Uk. ACM, May 2019. ISBN: 978-1-4503-5970-2. DOI: [10 . 1145/3290605 . 3300339](https://doi.org/10.1145/3290605.3300339).
- [319] Zachary Munn, Craig Lockwood, and Sandeep Moola. The development and use of evidence summaries for point of care information systems: a streamlined rapid review approach. *Worldviews on Evidence-Based Nursing*, 12(3):131–138, June 2015. ISSN: 1545-102X, 1741-6787. DOI: [10 . 1111/wvn.12094](https://doi.org/10.1111/wvn.12094).
- [320] Elizabeth L. Murnane, Tara G. Walker, Beck Tench, Stephen Voida, and Jaime Snyder. Personal informatics in interpersonal contexts: towards the design of technology that supports the social ecologies of long-term mental health management. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–27, Nov. 2018. ISSN: 2573-0142. DOI: [10 . 1145/3274396](https://doi.org/10.1145/3274396).
- [321] J. T. Nagurney. The accuracy and completeness of data collected by prospective and retrospective methods. *Academic Emergency Medicine*, 12(9):884–895, Sept. 2005. ISSN: 1069-6563. DOI: [10 . 1197 / j . aem . 2005 . 04 . 021](https://doi.org/10.1197/j.aem.2005.04.021).
- [322] Arvind Narayanan, Arunesh Mathur, Marshini Chetty, and Mihir Kshirsagar. Dark patterns: past, present, and future: the evolution of tricky user interfaces. *Queue*, 18(2):Pages 10:67–Pages 10:92, May 2020. ISSN: 1542-7730. DOI: [10 . 1145/3400899 . 3400901](https://doi.org/10.1145/3400899.3400901).
- [323] Simone Natale. *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test*. Oxford University Press, May 2021. ISBN: 978-0-19-008036-5. DOI: [10 . 1093/oso/9780190080365 . 003 . 0001](https://doi.org/10.1093/oso/9780190080365.003.0001).
- [324] Lei Nelissen and Mathias Funk. Rationalizing dark patterns: examining the process of designing privacy UX through speculative enactments, 2022. DOI: [10 . 57698/v16i1 . 05](https://doi.org/10.57698/v16i1.05).
- [325] Carman Neustaedter and A. J. Bernheim Brush. “linc-ing” the family: the participatory design of an inkable family calendar. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, 141–150, New York, NY, USA. Association for Computing Machinery, Apr. 2006. ISBN: 978-1-59593-372-0. DOI: [10 . 1145/1124772 . 1124796](https://doi.org/10.1145/1124772.1124796).

- [326] Carman Neustaedter and Phoebe Sengers. Autobiographical design in hci research: designing and learning through use-it-yourself. In *Proceedings of the Designing Interactive Systems Conference*, DIS '12, 514–523, Newcastle Upon Tyne, United Kingdom. Association for Computing Machinery, 2012. ISBN: 9781450312103. DOI: [10.1145/2317956.2318034](https://doi.org/10.1145/2317956.2318034). URL: <https://doi.org/10.1145/2317956.2318034>.
- [327] Jelena Nikolić. Biases in the decision-making process and possibilities of overcoming them. *Ekonomski horizonti*, 20(1):45–59, 2018. ISSN: 1450-863x. DOI: [10.5937/ekonhor1801045N](https://doi.org/10.5937/ekonhor1801045N).
- [328] Apichaya Nimkoompai. Risk analysis of encountering dark patterns of ux e-commerce applications affecting personal data. In *2022 6th International Conference on Information Technology (InCIT)*, 115–119, Nov. 2022. DOI: [10.1109/InCIT56086.2022.10067640](https://doi.org/10.1109/InCIT56086.2022.10067640).
- [329] Helen Fay Nissenbaum. *Privacy in context: technology, policy, and the integrity of social life*. Stanford Law Books, Stanford, Calif, 2010. ISBN: 978-0-8047-5236-7.
- [330] Robert Noggle. Manipulation, salience, and nudges. *Bioethics*, 32(3):164–170, Mar. 2018. ISSN: 02699702. DOI: [10.1111/bioe.12421](https://doi.org/10.1111/bioe.12421).
- [331] Robert Noggle. *The ethics of manipulation*. In *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta, editor. summer 2022 edition. Metaphysics Research Lab, Stanford University, summer 2022 edition, 2022. URL: <https://plato.stanford.edu/archives/sum2022/entries/ethics-manipulation/>.
- [332] Charlotte Nordmoen, Jack Armitage, Fabio Morreale, Rebecca Stewart, and Andrew McPherson. Making sense of sensors: discovery through craft practice with an open-ended sensor material. In *Proceedings of the 2019 on Designing Interactive Systems Conference*, 135–146, San Diego CA USA. ACM, June 2019. ISBN: 978-1-4503-5850-7. DOI: [10.1145/3322276.3322368](https://doi.org/10.1145/3322276.3322368).
- [333] Donald A Norman. *Emotional design: Why we love (or hate) everyday things*. Civitas Books, 2004.
- [334] Donald A. Norman. Signifiers, not affordances. *Interactions*:18–19, Nov. 2008. ISSN: 1072-5520, 1558-3449.
- [335] Donald A. Norman. *The design of everyday things*. Basic Books, New York, New York, revised and expanded edition edition, 2013. ISBN: 978-0-465-05065-9.
- [336] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. Dark patterns after the gdpr: scraping consent pop-ups and demonstrating their influence. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13, Apr. 2020. DOI: [10.1145/3313831.3376321](https://doi.org/10.1145/3313831.3376321). arXiv:2001.02479 [cs].
- [337] Thomas RV Nys and Bart Engelen. Judging nudging: answering the manipulation objection. *Political Studies*, 65(1):199–214, Mar. 2017. ISSN: 0032-3217, 1467-9248. DOI: [10.1177/0032321716629487](https://doi.org/10.1177/0032321716629487).
- [338] Tobias Nyström and Agnis Stibe. *When persuasive technology gets dark?* In *Information Systems*. Marinos Themistocleous, Maria Papadaki, and Muhammad Mustafa Kamal, editors. Volume 402. Lecture Notes in Business Information Processing. Springer International Publishing, Cham, 2020, 331–345. ISBN: 978-3-030-63395-0. DOI: [10.1007/978-3-030-63396-7_22](https://doi.org/10.1007/978-3-030-63396-7_22).
- [339] Ikechukwu Obi, Colin M Gray, Shruthi Sai Chivukula, Ja-Nae Duane, Janna Johns, Matthew Will, Ziqing Li, and Thomas Carlock. Let’s talk about socio-technical angst: tracing the history and evolution of dark patterns on twitter from 2010–2021:31, 2022. URL: <https://arxiv.org/abs/2207.10563>.
- [340] OECD. *Consumer Vulnerability in the Digital Age*, number 355 in OECD Digital Economy Papers. 2023, page 80.
- [341] Oecd. *Dark commercial patterns*, number 336 in OECD Digital Economy Papers. Oct. 2022, page 96.
- [342] Information Commissioner’s Office. *Overview of Data Protection Harms and the ICO’s Taxonomy*. Apr. 2022. URL: <https://ico.org.uk/media/about-the-ico/documents/4020144/overview-of-data-protection-harms-and-the-ico-taxonomy-v1-202204.pdf>.
- [343] Authority on Consumers and Markets (ACM). *Protection of the Online Consumer Boundaries of Online Persuasion*. The Netherlands, 2020.
- [344] Isabelle Oomen and Ronald Leenes. *Privacy risk perceptions and privacy protection strategies*. In *Policies and Research in Identity Management*. Elisabeth Leeuw, Simone Fischer-Hübner, Jimmy Tseng, and John Borking, editors. Volume 261. The International Federation for Information Processing. Springer US, Boston, MA, 2008, 121–138. ISBN: 978-0-387-77995-9. DOI: [10.1007/978-0-387-77996-6_10](https://doi.org/10.1007/978-0-387-77996-6_10).

- [345] M. O'Reilly and N. Kiyimba. *Advanced Qualitative Research: A Guide to Using Theory*. SAGE Publications, 2015. ISBN: 9781446273432.
- [346] BEUC. The European Consumer Organisation. Dark patterns and the eu consumer law acquis. recommendations for better enforcement and reform₂022, July 2022. URL: https://www.beuc.eu/sites/default/files/publications/beuc-x-2022-013_dark_patterns_paper.pdf.
- [347] Kentrell Owens, Johanna Gunawan, David Choffnes, Pardis Emami-Naeini, Tadayoshi Kohno, and Franziska Roesner. Exploring deceptive design patterns in voice interfaces. In *2022 European Symposium on Usable Security*, EuroUSEC '22, 64–78, Karlsruhe Germany. ACM, Sept. 2022. ISBN: 978-1-4503-9700-1. DOI: 10.1145/3549015.3554213.
- [348] Jennifer E. O'Brien, Kathryn B. Brewer, Lisa M. Jones, Jena Corkhum, and Cynthia Fraga Rizo. Rigor and respect: recruitment strategies for engaging vulnerable populations in research. *Journal of Interpersonal Violence*, 37(17–18):Np17052–np17072, Sept. 2022. ISSN: 1552-6518. DOI: 10.1177/08862605211023497.
- [349] Brian O'Neill. Ecological perspectives and children's use of the internet: exploring micro to macro level analysis. *Eesti Haridusteaduste Ajakiri. Estonian Journal of Education*, 3:32, Nov. 2015. DOI: 10.12697/eha.2015.3.2.02b.
- [350] Guido Palazzo, Franciska Krings, and Ulrich Hoffrage. Ethical blindness. *SSRN Electronic Journal*, 2013. ISSN: 1556-5068. DOI: 10.2139/ssrn.2212617.
- [351] Luci Pangrazio and Julian Sefton-Green. The social utility of 'data literacy'. *Learning, Media and Technology*, 45(2):208–220, Apr. 2020. ISSN: 1743-9884, 1743-9892. DOI: 10.1080/17439884.2020.1707223.
- [352] Luci Pangrazio and Neil Selwyn. "my data, my bad ...": young people's personal data understandings and (counter)practices. In *Proceedings of the 8th International Conference on Social Media & Society - #SMSociety17*, 1–5, Toronto, ON, Canada. ACM Press, 2017. ISBN: 978-1-4503-4847-8. DOI: 10.1145/3097286.3097338.
- [353] Luci Pangrazio and Neil Selwyn. 'personal data literacies': a critical literacies approach to enhancing understandings of personal digital data. *New Media & Society*, 21(2):419–437, Feb. 2019. ISSN: 1461-4448, 1461-7315. DOI: 10.1177/1461444818799523.
- [354] European Parliament and the Council. Regulation (eu) 2022 of the european parliament and of the council on a single market for digital services (digital services act) and amending directive 2000/31/ec₂016, 2016. DOI: 10.5040/9781782258674.
- [355] Article 29 Data Protection Working Party. *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*. Oct. 2017. URL: <https://ec.europa.eu/newsroom/article29/items/612053/en>.
- [356] Michael Quinn Patton. *Qualitative Research & Evaluation Methods: Integrating Theory and Practice*. SAGE Publications, Oct. 2014. ISBN: 978-1-4833-0145-7. Google-Books-ID: ovAkBQAAQBAJ.
- [357] Sebastiaan T.M. Peek, Eveline J.M. Wouters, Joost Van Hoof, Katrien G. Luijckx, Hennie R. Boeije, and Hubertus J.M. Vrijhoef. Factors influencing acceptance of technology for aging in place: a systematic review. *International Journal of Medical Informatics*, 83(4):235–248, Apr. 2014. DOI: 10.1016/j.ijmedinf.2014.01.004.
- [358] Eva PenzeyMoog. *Design for Safety*. URL: <https://abookapart.com/products/design-for-safety.html>.
- [359] Mark Pettigrew, Nason Maani, Luisa Pettigrew, Harry Rutter, and May Ci Van Schalkwyk. Dark nudges and sludge in big alcohol: behavioral economics, cognitive biases, and alcohol industry corporate social responsibility. *The Milbank Quarterly*, 98(4):1290–1328, 2020. ISSN: 1468-0009. DOI: 10.1111/1468-0009.12475.
- [360] James Pierce, Phoebe Sengers, Tad Hirsch, Tom Jenkins, William Gaver, and Carl DiSalvo. Expanding and refining design and criticality in hci. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2083–2092, Seoul Republic of Korea. ACM, Apr. 2015. ISBN: 978-1-4503-3145-6. DOI: 10.1145/2702123.2702438.
- [361] Laura R. Pina, Sang-Wha Sien, Teresa Ward, Jason C. Yip, Sean A. Munson, James Fogarty, and Julie A. Kientz. From personal informatics to family informatics: understanding family practices around health monitoring. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, 2300–2315, New York, NY, USA. Association for Computing Machinery, Feb. 2017. ISBN: 978-1-4503-4335-0. DOI: 10.1145/2998181.2998362.

- [362] Hanna F. Pitkin. *The attack of the blob: Hannah arendt's concept of the social*. University of Chicago Press, 1998. ISBN: 978-0-226-66990-8.
- [363] Thomas Poell, David Nieborg, and José Van Dijck. Platformisation. *Internet Policy Review*, 8(4), 2019. ISSN: 2197-6775. DOI: [10.14763/2019.4.1425](https://doi.org/10.14763/2019.4.1425).
- [364] Prisma statement. en-US. URL: <https://www.prisma-statement.org/prisma-2020-flow-diagram>.
- [365] Manuel Puppis. Media governance: a new concept for the analysis of media policy and regulation. *Communication, Culture & Critique*, 3(2):134-149, June 2010. ISSN: 17539129, 17539137. DOI: [10.1111/j.1753-9137.2010.01063.x](https://doi.org/10.1111/j.1753-9137.2010.01063.x).
- [366] Aditya Kumar Purohit, Kristoffer Bergram, Louis Barclay, Valéry Bezençon, and Adrian Holzer. Starving the newsfeed for social media detox: effects of strict and self-regulated facebook newsfeed diets. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1-16, Hamburg Germany. ACM, 2023. ISBN: 978-1-4503-9421-5. DOI: [10.1145/3544548.3581187](https://doi.org/10.1145/3544548.3581187).
- [367] Jenny Radesky, Alexis Hiniker, Caroline McLaren, Eliz Akgun, Alexandria Schaller, Heidi M. Weeks, Scott Campbell, and Ashley N. Gearhardt. Prevalence and characteristics of manipulative design in mobile applications used by children. *JAMA network open*, 5(6):e2217641, June 2022. ISSN: 2574-3805. DOI: [10.1001/jamanetworkopen.2022.17641](https://doi.org/10.1001/jamanetworkopen.2022.17641).
- [368] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. en, 2016. DOI: [10.5040/9781782258674](https://doi.org/10.5040/9781782258674). (Visited on 06/07/2021).
- [369] Karen Renaud, Cigdem Sengul, Kovila Coopamootoo, Bryan Clift, Jacqui Taylor, Mark Springett, and Ben Morrison. 'we're not that gullible!' revealing dark pattern mental models of 11-12 year-old scottish children. *ACM Transactions on Computer-Human Interaction*:3660342, Apr. 2024. ISSN: 1073-0516, 1557-7325. DOI: [10.1145/3660342](https://doi.org/10.1145/3660342).
- [370] Karen Renaud and Verena Zimmermann. Ethical guidelines for nudging in information security & privacy. *International Journal of Human-Computer Studies*, 120:22-35, Dec. 2018. ISSN: 10715819. DOI: [10.1016/j.ijhcs.2018.05.011](https://doi.org/10.1016/j.ijhcs.2018.05.011).
- [371] Orlando Ricciardi, Piero Maggi, and Francesco Di Nocera. Boredom makes me "nervous": fidgeting as a strategy for contrasting the lack of variety. *International Journal of Human Factors and Ergonomics*, 6:195, Jan. 2019. DOI: [10.1504/ijhfe.2019.104590](https://doi.org/10.1504/ijhfe.2019.104590).
- [372] Valeria Righi, Sergio Sayago, and Josep Blat. When we talk about older people in hci, who are we talking about? towards a 'turn to community' in the design of technologies for a growing ageing population. *International Journal of Human-Computer Studies*, 108:15-31, Dec. 2017. ISSN: 10715819. DOI: [10.1016/j.ijhcs.2017.06.005](https://doi.org/10.1016/j.ijhcs.2017.06.005).
- [373] Inger Roos. Methods of investigating critical incidents: a comparative review. *Journal of Service Research*, 4(3):193-204, Feb. 2002. ISSN: 1094-6705, 1552-7379. DOI: [10.1177/1094670502004003003](https://doi.org/10.1177/1094670502004003003).
- [374] Arianna Rossi and Kerstin Bongard-Blanchy. All in one stroke? intervention spaces for dark patterns. In *Position Paper at the Workshop "What Can CHI Do About Dark Patterns?"*, page 5, Online Virtual Conference (originally Yokohama, Japan), May 2021.
- [375] Arianna Rossi, Rachele Carli, Marietjie W. Botes, Angelica Fernandez, Anastasia Sergeeva, and Lorena Sánchez Chamorro. Who is vulnerable to deceptive design patterns? a transdisciplinary perspective on the multi-dimensional nature of digital vulnerability. *Computer Law & Security Review*, 55:106031, Nov. 2024. ISSN: 0267-3649. DOI: [10.1016/j.clsr.2024.106031](https://doi.org/10.1016/j.clsr.2024.106031).
- [376] Joel Rudinow. Manipulation. *Ethics*, 88(4):338-347, 1978. URL: <http://www.jstor.org/stable/2380239>.
- [377] Arianna Sala, Lorenzo Porcaro, and Emilia Gómez. Social media use and adolescents' mental health and well-being: an umbrella review. *Computers in Human Behavior Reports*, 14:100404, May 2024. ISSN: 24519588. DOI: [10.1016/j.chbr.2024.100404](https://doi.org/10.1016/j.chbr.2024.100404).

- [378] Johnny Saldaña. *The Coding Manual for Qualitative Researchers*. Sage, London, 2nd edition, 2013. ISBN: 978-1-4462-4736-5.
- [379] Lorena Sánchez Chamorro, Kerstin Bongard-Blanchy, and Vincent Koenig. Ethical tensions in UX design practice: exploring the fine line between persuasion and manipulation in online interfaces. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, Dis '23, 2408–2422, Pittsburgh PA USA. ACM, July 2023. ISBN: 978-1-4503-9893-0. DOI: [10.1145/3563657.3596013](https://doi.org/10.1145/3563657.3596013).
- [380] Lorena Sánchez Chamorro, Kerstin Bongard-Blanchy, and Vincent Koenig. Justice in interaction design: preventing manipulation in interfaces. *Dreaming Disability Justice Workshop at Human Factors in Computer Systems Conference 2021*. New Orleans, LA. (Hybrid conference), (arXiv:2204.06821), May 2022. URL: <http://arxiv.org/abs/2204.06821>. arXiv:2204.06821 [cs].
- [381] Lorena Sánchez Chamorro and Carine Lallemand. Towards a second wave of manipulative design research: methodological challenges of studying the effects of manipulative designs on users. In Honolulu (HI) (Hybrid workshop), May 2024.
- [382] Lorena Sánchez Chamorro, Carine Lallemand, and Colin M. Gray. "My mother told me these things are always fake" - understanding teenagers' experiences with manipulative designs. In *Designing Interactive Systems Conference*, 1469–1482, IT University of Copenhagen Denmark. ACM, July 2024. DOI: [10.1145/3643834.3660704](https://doi.org/10.1145/3643834.3660704).
- [383] Lorena Sánchez Chamorro, Romain Toebosch, and Carine Lallemand. Manipulative design and older adults: co-creating magic machines to understand experiences of online manipulation. In *Designing Interactive Systems Conference*, 668–684, IT University of Copenhagen Denmark. ACM, July 2024. DOI: [10.1145/3643834.3661513](https://doi.org/10.1145/3643834.3661513).
- [384] Hauke Sandhaus. Promoting bright patterns. (arXiv:2304.01157), Apr. 2023. URL: <http://arxiv.org/abs/2304.01157>.
- [385] Shruti Sannon and Andrea Forte. Privacy research with marginalized groups: what we know, what's needed, and what's next. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2):1–33, Nov. 2022. ISSN: 2573-0142. DOI: [10.1145/3555556](https://doi.org/10.1145/3555556).
- [386] Cristiana Santos, Viktorija Morozovaite, and Silvia De Conca. No harm no foul: how harms caused by dark patterns are conceptualised and tackled under eu data protection, consumer and competition laws, 2024. DOI: [10.2139/ssrn.4877439](https://doi.org/10.2139/ssrn.4877439).
- [387] Cristiana Santos, Arianna Rossi, Lorena Sánchez Chamorro, Kerstin Bongard-Blanchy, and Ruba Abu-Salma. Cookie banners, what's the purpose?: analyzing cookie banner text through a legal lens. In *Proceedings of the 20th Workshop on Workshop on Privacy in the Electronic Society*, 187–194, Virtual Event Republic of Korea. ACM, 2021. ISBN: 978-1-4503-8527-5. DOI: [10.1145/3463676.3485611](https://doi.org/10.1145/3463676.3485611).
- [388] René Schäfer, Paul Miles Preuschoff, and Jan Borchers. Investigating visual countermeasures against dark patterns in user interfaces. In *Mensch und Computer 2023*, MuC '23, 161–172, Rapperswil Switzerland. ACM, Sept. 2023. ISBN: 9798400707711. DOI: [10.1145/3603555.3603563](https://doi.org/10.1145/3603555.3603563).
- [389] René Schäfer, Paul Miles Preuschoff, René Röpke, Sarah Sahabi, and Jan Borchers. Fighting malicious designs: towards visual countermeasures against dark patterns. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–13, Honolulu HI USA. ACM, May 2024. ISBN: 9798400703300. DOI: [10.1145/3613904.3642661](https://doi.org/10.1145/3613904.3642661).
- [390] Brennan Schaffner, Neha A. Lingareddy, and Marshini Chetty. Understanding account deletion and relevant dark patterns on social media. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2):1–43, Nov. 2022. ISSN: 2573-0142. DOI: [10.1145/3555142](https://doi.org/10.1145/3555142).
- [391] A. J. Scheerder, A. J. A. M. van Deursen, and J. A. G. M van Dijk. Taking advantage of the internet: a qualitative analysis to explain why educational background is decisive in gaining positive outcomes. *Poetics*, 80:101426, June 2020. ISSN: 0304-422x. DOI: [10.1016/j.poetic.2019.101426](https://doi.org/10.1016/j.poetic.2019.101426).
- [392] Sarita Schoenebeck and Lindsay Blackwell. Reimagining social media governance: harm, accountability, and repair. *SSRN Electronic Journal*, 2021. ISSN: 1556-5068. DOI: [10.2139/ssrn.3895779](https://doi.org/10.2139/ssrn.3895779).
- [393] Katie Seaborn and Weichen Joe Chang. Another subtle pattern: examining demographic biases in dark patterns and deceptive design research. In Honolulu (HI) (Hybrid workshop), May 2024.

- [394] Katie Seaborn, Tatsuya Itagaki, Mizuki Watanabe, Yijia Wang, Ping Geng, Takao Fujii, Yuto Mandai, Miu Kojima, and Suzuka Yoshida. Deceptive, disruptive, no big deal: japanese people react to simulated dark commercial patterns. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 1–8, Honolulu HI USA. ACM, May 2024. ISBN: 9798400703317. DOI: [10.1145/3613905.3651099](https://doi.org/10.1145/3613905.3651099).
- [395] Amartya Sen. *Freedom and the foundations of justice*. In *Development as Freedom*. Oxford University Press, 1999, 54–87.
- [396] Anastasia Sergeeva, Kerstin Bongard-Blanchy, Verena Distler, and Vincent Koenig. The perceived influence of e-shopping cues on customers' buying decisions. In *Proceedings of the XI Latin American Conference on Human Computer Interaction*, 1–8, Puebla Mexico. ACM, Oct. 2023. ISBN: 9798400716577. DOI: [10.1145/3630970.3630980](https://doi.org/10.1145/3630970.3630980).
- [397] William Seymour and Max Van Kleek. Exploring interactions between trust, anthropomorphism, and relationship development in voice assistants. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2):371:1–371:16, Oct. 2021. DOI: [10.1145/3479515](https://doi.org/10.1145/3479515).
- [398] Abdolreza Shaghghi, Raj S Bhopal, and Aziz Sheikh. Approaches to recruiting 'hard-to-reach' populations into re- search: a review of the literature. *Health Promotion Perspectives*, 1(2).
- [399] Naveen Shamsudhin and Fabrice Jotterand. *Social robots and dark patterns: where does persuasion end and deception begin?* In *Artificial Intelligence in Brain and Mental Health: Philosophical, Ethical & Policy Issues*. Fabrice Jotterand and Marcello Ienca, editors. Advances in Neuroethics. Springer International Publishing, Cham, 2021, 89–110. ISBN: 978-3-030-74187-7. DOI: [10.1007/978-3-030-74188-4_7](https://doi.org/10.1007/978-3-030-74188-4_7).
- [400] Sandeep Sharma and Ishita Sharma. Dark patterns in a bright world: an analysis of the indian consumer legal architecture. *International Journal on Consumer Law and Practice*, 11(1), Nov. 2023. ISSN: 2347-2731. URL: <https://repository.nls.ac.in/ijclp/vol11/iss1/7>.
- [401] Ashley Sheil, Gunes Acar, Hanna Schraffenberger, Raphaël Gellert, and David Malone. Staying at the roach motel: cross-country analysis of manipulative subscription and cancellation flows. (arXiv:2309.17145), Dec. 2023. URL: <http://arxiv.org/abs/2309.17145>. arXiv:2309.17145 [cs].
- [402] Stephen Tsung-Han Sher and Norman Makoto Su. From screens to projector, wall, and tvs: conceptualizing livestreams as design material for direct and indirect viewership experiences. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1):1–22, Apr. 2023. ISSN: 2573-0142. DOI: [10.1145/3579489](https://doi.org/10.1145/3579489).
- [403] Zewei Shi, Ruoxi Sun, Jieshan Chen, Jiamou Sun, and Minhui Xue. The invisible game on the internet: a case study of decoding deceptive patterns. In *Companion Proceedings of the ACM Web Conference 2024*, 521–524, Singapore Singapore. ACM, May 2024. ISBN: 9798400701726. DOI: [10.1145/3589335.3651571](https://doi.org/10.1145/3589335.3651571).
- [404] Ruth Shillair, Shelia R. Cotten, Hsin-Yi Sandy Tsai, Saleem Alhabash, Robert LaRose, and Nora J. Rifon. On-line safety begins with you and me: convincing internet users to protect themselves. *Computers in Human Behavior*, 48:199–207, 2015. ISSN: 07475632. DOI: [10.1016/j.chb.2015.01.046](https://doi.org/10.1016/j.chb.2015.01.046).
- [405] Paolo Siciliani, Christine Riefa, and Harriet Gamper. *Consumer Theories of Harm: An Economic Approach to Consumer Law Enforcement and Policy Making*. June 2019. ISBN: 978-1-5099-1685-6. DOI: [10.5040/9781509916887](https://doi.org/10.5040/9781509916887).
- [406] Adam Silver. Why it's time to update our language about bad design patterns by amy hupe, content designer. URL: <https://amyhupe.co.uk/articles/changing-our-language-on-bad-patterns/>.
- [407] H. A. Simon. Rational choice and the structure of the environment. *Psychological Review*, 63(2):129–138, 1956. ISSN: 1939-1471, 0033-295x. DOI: [10.1037/h0042769](https://doi.org/10.1037/h0042769).
- [408] Jaisie Sin and Cosmin Munteanu. A preliminary investigation of the role of anthropomorphism in designing telehealth bots for older adults. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–6, Glasgow Scotland Uk. ACM, May 2019. ISBN: 978-1-4503-5971-9. DOI: [10.1145/3290607.3312941](https://doi.org/10.1145/3290607.3312941).
- [409] Ray Sin, Ted Harris, Simon Nilsson, and Talia Beck. Dark patterns in online shopping: do they work and can nudges help mitigate impulse buying? *Behavioural Public Policy*:1–27, May 2022. ISSN: 2398-063x, 2398-0648. DOI: [10.1017/bpp.2022.11](https://doi.org/10.1017/bpp.2022.11).
- [410] caroline sinders. What's in a name?, June 2022. URL: <https://medium.com/@carolinesinders/whats-in-a-name-unpacking-dark-patterns-versus-deceptive-design-e96068627ec4>.

- [411] Ashwin Singh, Arvinth Arun, Ayushi Jain, Pooja Desur, Pulak Malhotra, Duen Horng Chau, and Ponnurangam Kumaraguru. Erasing labor with labor: dark patterns and lockstep behaviors on google play. In *Proceedings of the 33rd ACM Conference on Hypertext and Social Media*, 186–191, June 2022. DOI: [10.1145/3511095.3536368](https://doi.org/10.1145/3511095.3536368). arXiv:2202.04561 [cs].
- [412] Mahender Singh and Anurag Vidyarthi. *Digital Literacy towards Dark Pattern among Research Scholars of Babasaheb Bhimrao Ambedkar Central University, Lucknow, Uttar Pradesh: A Study*. Apr. 2023.
- [413] Michael Skirpan, Nathan Beard, Srinjita Bhaduri, Casey Fiesler, and Tom Yeh. Ethics education in context: a case study of novel ethics activities for the cs classroom. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, 940–945, Baltimore Maryland USA. ACM, Feb. 2018. ISBN: 978-1-4503-5103-4. DOI: [10.1145/3159450.3159573](https://doi.org/10.1145/3159450.3159573).
- [414] Mira Slavova. Towards applications of capability sensitive design of technologies. *IFIP'14*, Apr. 2013.
- [415] Julia Slupska and Leonie Maria Tanczer. *Threat modeling intimate partner violence: tech abuse as a cybersecurity challenge in the internet of things*. In *The Emerald International Handbook of Technology-Facilitated Violence and Abuse*. Jane Bailey, Asher Flynn, and Nicola Henry, editors. Emerald Publishing Limited, June 2021, 663–688. ISBN: 978-1-83982-849-2. DOI: [10.1108/978-1-83982-848-520211049](https://doi.org/10.1108/978-1-83982-848-520211049).
- [416] Nathalie A. Smuha, Emma Ahmed-Rengers, Adam Harkens, Wenlong Li, James MacLaren, Riccardo Piselli, and Karen Yeung. How the eu can achieve legally trustworthy ai: a response to the european commission's proposal for an artificial intelligence act. (3899991), Aug. 2021. DOI: [10.2139/ssrn.3899991](https://doi.org/10.2139/ssrn.3899991).
- [417] Ryan Snelgrove and Mark E. Havitz. Looking back in time: the pitfalls and potential of retrospective methods in leisure studies. *Leisure Sciences*, 32(4):337–351, 2010. ISSN: 0149-0400, 1521-0588. DOI: [10.1080/01490400.2010.488199](https://doi.org/10.1080/01490400.2010.488199).
- [418] Robert Soden, Austin Toombs, and Michaelanne Thomas. Evaluating interpretive research in HCI. *Interactions*, 31(1):38–42, Jan. 2024. ISSN: 1072-5520, 1558-3449. DOI: [10.1145/3633200](https://doi.org/10.1145/3633200).
- [419] Than Htut Soe, Oda Elise Nordberg, Frode Guribye, and Marija Slavkovic. Circumvention by design - dark patterns in cookie consent for online news outlets. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, 1–12, Tallinn Estonia. ACM, Oct. 2020. ISBN: 978-1-4503-7579-5. DOI: [10.1145/3419249.3420132](https://doi.org/10.1145/3419249.3420132). arXiv: 2006.13985.
- [420] Carla Sousa and Ana Oliveira. The dark side of fun: understanding dark patterns and literacy needs in early childhood mobile gaming. *European Conference on Games Based Learning*, 17(1):599–610, Sept. 2023. ISSN: 2049-100x, 2049-0992. DOI: [10.34190/ecgb1.17.1.1656](https://doi.org/10.34190/ecgb1.17.1.1656).
- [421] Shaun B. Spencer. The problem of online manipulation. *SSRN Electronic Journal*, 2019. ISSN: 1556-5068. DOI: [10.2139/ssrn.3341653](https://doi.org/10.2139/ssrn.3341653).
- [422] Bernd Stahl, Rowena Rodrigues, Nicole Santiago, and Kevin Macnish. A european agency for artificial intelligence: protecting fundamental rights and ethical values. *Computer Law & Security Review*, 45:105661, July 2022. DOI: [10.1016/j.clsr.2022.105661](https://doi.org/10.1016/j.clsr.2022.105661).
- [423] Renée E. Stalmeijer, Megan E. L. Brown, and Bridget C. O'Brien. How to discuss transferability of qualitative research in health professions education. en. *The Clinical Teacher*, 21(6):e13762, Dec. 2024. ISSN: 1743-4971, 1743-498X. DOI: [10.1111/tct.13762](https://doi.org/10.1111/tct.13762).
- [424] Susan Leigh Star. The ethnography of infrastructure. *American Behavioral Scientist*, 43(3):377–391, Dec. 1999.
- [425] Jeremiah Still and Mary Still. Influence of visual salience on webpage product searches. *ACM Trans. Appl. Percept.*, 16(1):3:1–3:11, Feb. 2019. ISSN: 1544-3558. DOI: [10.1145/3301413](https://doi.org/10.1145/3301413).
- [426] Alina Stöver, Nina Gerber, Christin Cornel, Mona Henz, Karola Marky, Verena Zimmermann, and Joachim Vogt. Website operators are not the enemy either - analyzing options for creating cookie consent notices without dark patterns. *September*:9, 2022.
- [427] Joanna Strycharz, Guda Van Noort, Edith Smit, and Natali Helberger. Protective behavior against personalized ads: motivation to turn personalization off. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 13(2), 2019. ISSN: 1802-7962. DOI: [10.5817/cp2019-2-1](https://doi.org/10.5817/cp2019-2-1).

- [428] Daniel Susser and Vincent Grimaldi. Measuring automated influence: between empirical evidence and ethical values. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 242–253, Virtual Event USA. ACM, July 2021. ISBN: 978-1-4503-8473-5. DOI: [10.1145/3461702.3462532](https://doi.org/10.1145/3461702.3462532).
- [429] Daniel Susser, Beate Roessler, and Helen F. Nissenbaum. Online manipulation: hidden influences in a digital world. *SSRN Electronic Journal*, 2018. ISSN: 1556-5068. DOI: [10.2139/ssrn.3306006](https://doi.org/10.2139/ssrn.3306006).
- [430] Mohammad Tahaei and Kami Vaniea. “developers are responsible”: what ad networks tell developers about privacy. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–11, Yokohama Japan. ACM, May 2021. ISBN: 978-1-4503-8095-9. DOI: [10.1145/3411763.3451805](https://doi.org/10.1145/3411763.3451805).
- [431] Richard Thaler and Cass R. Sunstein. *Introduction*. In *Nudge. Improving Decisions About Health, Wealth, and Happiness*. Penguin Books, 2009, 1–17.
- [432] Richard H. Thaler, Cass R. Sunstein, and John P. Balz. *Chapter 25. Choice Architecture*. Princeton University Press, Jan. 2013. ISBN: 978-1-4008-4534-7. URL: <https://www.degruyter.com/document/doi/10.1515/9781400845347-029/html>.
- [433] Jelmer Tiemessen, Hanna Schraffenberger, and Gunes Acar. The time is ticking: the effect of limited time discounts on consumers’ buying behavior and experience. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–11, Hamburg Germany. ACM, Apr. 2023. ISBN: 978-1-4503-9422-2. DOI: [10.1145/3544549.3585735](https://doi.org/10.1145/3544549.3585735).
- [434] Ingvar Tjostheim, Vanessa Ayres-Pereira, Chris Wales, Angela Manna, and Simon Egenfeldt-Nielsen. Dark pattern: a serious game for learning about the dangers of sharing data. *European Conference on Games Based Learning*, 16(1):774–783, Sept. 2022. ISSN: 2049-100x, 2049-0992. DOI: [10.34190/ecgbl.16.1.872](https://doi.org/10.34190/ecgbl.16.1.872).
- [435] Austin Toombs, Shad Gross, Shaowen Bardzell, and Jeffrey Bardzell. From empathy to care: a feminist care ethics perspective on long-term researcher–participant relations. *Interacting with Computers*, 29(1):45–57, Jan. 2017. ISSN: 0953-5438, 1873-7951. DOI: [10.1093/iwc/iww010](https://doi.org/10.1093/iwc/iww010).
- [436] Michael Toth, Nataliia Bielova, and Vincent Roca. On dark patterns and manipulation of website publishers by cmps. *Proceedings on Privacy Enhancing Technologies*, 2022(3):478–497, July 2022. ISSN: 2299-0984. DOI: [10.56553/popets-2022-0082](https://doi.org/10.56553/popets-2022-0082).
- [437] Nynke Tromp, Paul Hekkert, and Peter-Paul Verbeek. Design for socially responsible behavior: a classification of influence based on intended user experience. *Design Issues*, 27(3):3–19, July 2011. ISSN: 0747-9360, 1531-4790. DOI: [10.1162/DESI_a_00087](https://doi.org/10.1162/DESI_a_00087).
- [438] Eric Tsetsi and Stephen Rains. Smartphone internet access and use: extending the digital divide and usage gap. *Mobile Media & Communication*, 5:205015791770832, June 2017. DOI: [10.1177/2050157917708329](https://doi.org/10.1177/2050157917708329).
- [439] Reha Tuncer, Anastasia Sergeeva, Kerstin Bongard-Blanchy, Verena Distler, Sophie Doublet, and Vincent Koenig. Running out of time(rs): effects of scarcity cues on perceived task load, perceived benevolence and user experience on e-commerce sites. *Behaviour & Information Technology*:1–19, Sept. 2023. ISSN: 0144-929x, 1362-3001. DOI: [10.1080/0144929x.2023.2242966](https://doi.org/10.1080/0144929x.2023.2242966).
- [440] Laia Turmo Vidal, José Manuel Vega-Cebrián, María Concepción Valdez Gastelum, Elena Márquez Segura, Judith Ley-Flores, Joaquin R. Diaz Duran, and Ana Tajadura-Jiménez. Body sensations as design material: an approach to design sensory technology for altering body perception. In *Designing Interactive Systems Conference*, 2545–2561, IT University of Copenhagen Denmark. ACM, July 2024. DOI: [10.1145/3643834.3660701](https://doi.org/10.1145/3643834.3660701).
- [441] Sweden Umeå University Digital Media Lab, Thomas Mejtoft, Erik Frångsmyr, Ulrik Söderström, Sweden Umeå University Digital Media Lab, Ole Norberg, and Sweden Umeå University Digital Media Lab. Deceptive design: cookie consent and manipulative patterns. In *34th Bled eConference Digital Support from Crisis to Progressive Change: Conference Proceedings*, 393–404. University of Maribor Press, 2021. ISBN: 978-961-286-485-9. DOI: [10.18690/978-961-286-485-9.29](https://doi.org/10.18690/978-961-286-485-9.29).
- [442] Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. (un)informed consent: studying gdpr consent notices in the field. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, 973–990, London United Kingdom. ACM, Nov. 2019. ISBN: 978-1-4503-6747-9. DOI: [10.1145/3319535.3354212](https://doi.org/10.1145/3319535.3354212). arXiv: 1909.02638.

- [443] Alexander J. A. M. van Deursen and Ellen J. Helsper. The third-level digital divide: who benefits most from being online? In Laura Robinson, Shelia R. Cotten, Jeremy Schulz, Timothy M. Hale, and Apryl Williams, editors, *Studies in Media and Communications*, volume 10, 29–52. Emerald Group Publishing Limited, Dec. 2015. ISBN: 978-1-78560-381-5. DOI: [10.1108/s2050-206020150000010002](https://doi.org/10.1108/s2050-206020150000010002).
- [444] Alexander J.A.M. van Deursen, Ellen J. Helsper, and Rebecca Eynon. Development and validation of the internet skills scale (iss). *Information, Communication & Society*, 19(6):804–823, June 2016. ISSN: 1369-118X, 1468-4462. DOI: [10.1080/1369118X.2015.1078834](https://doi.org/10.1080/1369118X.2015.1078834).
- [445] Jan van Dijk and Kenneth Hacker. The digital divide as a complex and dynamic phenomenon. *The Information Society*, 19(4):315–326, Sept. 2003. ISSN: 0197-2243, 1087-6537. DOI: [10.1080/01972240309487](https://doi.org/10.1080/01972240309487).
- [446] Jan van Dijk and Alexander J. A. M. van Deursen. *Digital Skills*. Palgrave Macmillan US, New York, 2014. ISBN: 978-1349-49370-8. DOI: [10.1057/9781137437037](https://doi.org/10.1057/9781137437037).
- [447] Christof van Nimwegen and Jesse de Wit. *Shopping in the dark: effects of platform choice on dark pattern recognition*. In *Human-Computer Interaction. User Experience and Behavior*. Masaaki Kurosu, editor. Volume 13304. Lecture Notes in Computer Science. Springer International Publishing, Cham, 2022, 462–475. ISBN: 978-3-031-05411-2. DOI: [10.1007/978-3-031-05412-9_32](https://doi.org/10.1007/978-3-031-05412-9_32).
- [448] Evalina van Wijk and Tracie Harrison. Managing ethical problems in qualitative research involving vulnerable populations, using a pilot study. *International Journal of Qualitative Methods*, 12(1):570–586, Feb. 2013. ISSN: 1609-4069, 1609-4069. DOI: [10.1177/160940691301200130](https://doi.org/10.1177/160940691301200130).
- [449] Alexander Jam Van Deursen and Ellen J Helsper. A nuanced understanding of internet use and non-use among the elderly. *European Journal of Communication*, 30(2):171–187, Apr. 2015. ISSN: 0267-3231, 1460-3705. DOI: [10.1177/0267323115578059](https://doi.org/10.1177/0267323115578059).
- [450] Eduardo Veas, Erick Mendez, Steven Feiner, and Dieter Schmalstieg. *Directing attention and influencing memory with visual saliency modulation*. May 2011, page 1480. DOI: [10.1145/1978942.1979158](https://doi.org/10.1145/1978942.1979158).
- [451] Peter-Paul Verbeek. Materializing morality: design ethics and technological mediation. *Science, Technology, & Human Values*, 31(3):361–380, 2006.
- [452] John Vines, Mark Blythe, Stephen Lindsay, Paul Dunphy, Andrew Monk, and Patrick Olivier. Questionable concepts: critique as resource for designing with eighty somethings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1169–1178, Austin Texas USA. ACM, May 2012. ISBN: 978-1-4503-1015-4. DOI: [10.1145/2207676.2208567](https://doi.org/10.1145/2207676.2208567).
- [453] Christian Voigt, Stephan Schlögl, and Aleksander Groth. *Dark patterns in online shopping: of sneaky tricks, perceived annoyance and respective brand trust*. In *HCI in Business, Government and Organizations*. Fiona Fui-Hoon Nah and Keng Siau, editors. Volume 12783. Lecture Notes in Computer Science. Springer International Publishing, Cham, 2021, 143–155. ISBN: 978-3-030-77749-4. DOI: [10.1007/978-3-030-77750-0_10](https://doi.org/10.1007/978-3-030-77750-0_10).
- [454] Toni Waefler and Ute Schmid. Explainability is not enough: requirements for human-ai- partnership in complex socio-technical systems, Mar. 2021. DOI: [10.34190/eair.20.007](https://doi.org/10.34190/eair.20.007).
- [455] Ari Ezra Waldman. Cognitive biases, dark patterns, and the ‘privacy paradox’. *Current Opinion in Psychology*, 31:105–109, 2020. ISSN: 2352250x. DOI: [10.1016/j.copsyc.2019.08.025](https://doi.org/10.1016/j.copsyc.2019.08.025).
- [456] Xian Wang, Lik-Hang Lee, Carlos Bermejo Fernandez, and Pan Hui. The dark side of augmented reality: exploring manipulative designs in ar. *International Journal of Human-Computer Interaction*, 40(13):3449–3464, July 2024. ISSN: 1044-7318, 1532-7590. DOI: [10.1080/10447318.2023.2188799](https://doi.org/10.1080/10447318.2023.2188799). arXiv:2303.02843 [cs].
- [457] Johanna I. Westbrook, Enrico W. Coiera, A. Sophie Gosling, and Jeffrey Braithwaite. Critical incidents and journey mapping as techniques to evaluate the impact of online evidence retrieval systems on health care delivery and patient outcomes. *International Journal of Medical Informatics. Connecting Medical Informatics and Bio-Informatics - MIE 2005*, 76(2):234–245, Feb. 2007. ISSN: 1386-5056. DOI: [10.1016/j.ijmedinf.2006.03.006](https://doi.org/10.1016/j.ijmedinf.2006.03.006).
- [458] Fiona Westin and Sonia Chiasson. Opt out of privacy or “go home”: understanding reluctant privacy behaviours through the fomo-centric design paradigm. In *Proceedings of the New Security Paradigms Workshop*, 57–67, San Carlos Costa Rica. ACM, Sept. 2019. ISBN: 978-1-4503-7647-1. DOI: [10.1145/3368860.3368865](https://doi.org/10.1145/3368860.3368865).

- [459] T. M. Wilkinson. Nudging and manipulation. *Political Studies*, 61(2):341–355, June 2013. ISSN: 0032-3217, 1467-9248. DOI: [10.1111/j.1467-9248.2012.00974.x](https://doi.org/10.1111/j.1467-9248.2012.00974.x).
- [460] Langdon Winner. Do artifacts have politics? *Daedalus*, 109(1):121–136, 1980.
- [461] Pamela Wisniewski, Jessica Vitak, and Heidi Hartikainen. *Privacy in adolescence*. In *Modern Socio-Technical Perspectives on Privacy*. Springer International Publishing, Cham, 2022. ISBN: 978-3-030-82785-4. DOI: [10.1007/978-3-030-82786-1](https://doi.org/10.1007/978-3-030-82786-1).
- [462] Richmond Y. Wong. Tactics of soft resistance in user experience professionals' values work. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–28, Oct. 2021. ISSN: 2573-0142. DOI: [10.1145/3479499](https://doi.org/10.1145/3479499).
- [463] Qunfang Wu, Yisi Sang, Dakuo Wang, and Zhicong Lu. Malicious selling strategies in livestream e-commerce: a case study of alibaba's taobao and bytedance's tiktok. *ACM Trans. Comput.-Hum. Interact.*, 30(3):35:1–35:29, June 2023. ISSN: 1073-0516. DOI: [10.1145/3577199](https://doi.org/10.1145/3577199).
- [464] Michaela Wänke. Primes as hidden persuaders. *Current Opinion in Psychology*. Social priming, 12:63–66, Dec. 2016. ISSN: 2352-250X. DOI: [10.1016/j.copsyc.2016.05.004](https://doi.org/10.1016/j.copsyc.2016.05.004).
- [465] Boming Xia, Qinghua Lu, Harsha Perera, Liming Zhu, Zhenchang Xing, Yue Liu, and Jon Whittle. *Towards Concrete and Connected AI Risk Assessment (C 2 AIRA): A Systematic Mapping Study*. May 2023, page 116. DOI: [10.1109/cain58948.2023.00027](https://doi.org/10.1109/cain58948.2023.00027).
- [466] Haiyan Xue and Pieter M.A. Desmet. Researcher introspection for experience-driven design research. *Design Studies*, 63:37–64, July 2019. ISSN: 0142694x. DOI: [10.1016/j.destud.2019.03.001](https://doi.org/10.1016/j.destud.2019.03.001).
- [467] Karen Yeung. 'hypernudge': big data as a mode of regulation by design. *Information, Communication & Society*, 20(1):118–136, Jan. 2017. ISSN: 1369-118x. DOI: [10.1080/1369118x.2016.1186713](https://doi.org/10.1080/1369118x.2016.1186713).
- [468] Youjae Yi. Contextual priming effects in print advertisements: the moderating role of prior knowledge. *Journal of Advertising*, 22(1):1–10, 1993. ISSN: 0091-3367.
- [469] Chan Yoo. Unconscious processing of web advertising: effects on implicit memory, attitude toward the brand, and consideration set. *Journal of Interactive Marketing*, 22:2–18, Mar. 2008. DOI: [10.1002/dir.20110](https://doi.org/10.1002/dir.20110).
- [470] Selman Yüçetürk, Mohammad Obaid, and Asım Evren Yantaç. Probing human-soundscape interaction using observational user experience methods. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, NordiCHI '16, 1–4, New York, NY, USA. Association for Computing Machinery, Oct. 2016. ISBN: 978-1-4503-4763-1. DOI: [10.1145/2971485.2971495](https://doi.org/10.1145/2971485.2971495).
- [471] Jennifer Preece Yvonne Rogers Helen Sharp. *Interaction Design: Beyond Human-Computer Interaction*. Praise for Interaction Design. John Wiley & Sons, 6th edition, 2023. ISBN: 9781119901099; 111990109X.
- [472] Amit Zac, Yun-Chun Huang, Amédée von Moltke, Christopher Decker, and Ariel Ezrachi. Dark patterns and consumer vulnerability, Aug. 2023. URL: <http://dx.doi.org/10.2139/ssrn.4547964>.
- [473] José P Zagal, Staffan Björk, and Chris Lewis. Dark patterns in the design of games. *Foundations of Digital Games 2013*:8, 2013.
- [474] Federica Zanca, Caterina Brusasco, Filippo Pesapane, Zuzanna Kwade, Ruth Beckers, and Michele Avanzo. Regulatory aspects of the use of artificial intelligence medical software. eng. *Seminars in Radiation Oncology*, 32(4):432–441, Oct. 2022. ISSN: 1532-9461. DOI: [10.1016/j.semradonc.2022.06.012](https://doi.org/10.1016/j.semradonc.2022.06.012).
- [475] David Zendle, Paul Cairns, Herbie Barnett, and Cade McCall. Paying for loot boxes is linked to problem gambling, regardless of specific features like cash-out and pay-to-win. *Computers in Human Behavior*, 102:181–191, Jan. 2020. ISSN: 07475632. DOI: [10.1016/j.chb.2019.07.003](https://doi.org/10.1016/j.chb.2019.07.003).
- [476] Leah Zhang-Kennedy, Maxwell Keleher, and Michaela Valiquette. Navigating the gray: design practitioners' perceptions toward the implementation of privacy dark patterns. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW1):1–26, Apr. 2024. ISSN: 2573-0142. DOI: [10.1145/3637374](https://doi.org/10.1145/3637374).
- [477] Wei Zhao, Ryan M Kelly, Melissa J Rogerson, and Jenny Waycott. Older adults imagining future technologies in participatory design workshops: supporting continuity in the pursuit of meaningful activities. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. ACM, 2024.

- [478] Luigi Zingales, Fiona Scott Morton, and Guy Rolnik. Stigler committee on digital platforms. URL: <https://www.chicagobooth.edu/-/media/research/stigler/pdfs/digital-platforms-committee-report---stigler-center.pdf>.
- [479] Shoshana Zuboff. Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1):75–89, Mar. 2015. ISSN: 0268-3962, 1466-4437. DOI: 10.1057/jit.2015.5.
- [480] Shoshana Zuboff. The right to sanctuary. In *The Age of Surveillance Capitalism*, 475–494. Profile Books, 2019.

Chapter 13

Appendix

These appendices include some supplementary material used in the different chapters of the dissertation, and that belong to the different studies

13.1 Chapter 5

13.1.1 Data Analysis

Here I present a non-exhaustive example of codes used to create the themes and overarching themes building on Braun and Clarke [51] thematic analysis methodology.

Overarching theme	Themes	Codes
Influencing is an exchange	Convincing users with reasons Exchange between users and designers Understanding how mechanisms work on users Friction and stickiness	Rational arguments You give me your email, and I give you something in return Added value for the user through the newsletter Tangible incentives Intangible incentives Analytics Need to test Elements to catch attention Permanent reminders Non-intrusive elements

Overarching theme	Themes	Codes
Conditions of manipulation and guiding principles	Conditions of manipulation	Designers know how to manipulate The business is not ethical Knowing the classic DP: coercion and deception
	Impacts	An emotion to communicate is the beginning of manipulation Psychological aspects of users Environmental aspect of impacts The main impact is getting subscribers: to comply with our mission They can avoid it: close the webpage, do something active
	Autonomy	I am not pushing or forcing the user I am not lying, not tricking The information needs to be complete
	Trust and transparency	The information needs to be complete Human-readable designs We need trust to get the e-mails
	Usability	Designers expect that user behaves in a certain way Designers expect that user behaves in a certain way Balance between what they know and the information This is very common mechanism, they are used to
	Users vs user experience first	Pretty direct and obvious, they know it I don't want to transmit negative feelings I don't want to bother the user It's user first

Overarching theme	Themes	Codes
Responsibilities and hurdles	Design responsibility is shared	Designers also teach Own values matter Designers rely on rules in the company
	Imbalanced power towards business	If the company want us to do DP we will do it I can choose my company

13.2 Chapter 8

13.2.1 Study Materials

Here I present some materials to provide information in the recruitment and debriefing.

13.2.2 Recruitment and debriefing materials

Here I show an example of the poster we used to give information to participants before the study (Fig. 13.1) and the poster sent to NGOs to remind them about the debriefing information (Fig. 13.2).



Figure 13.1: Example of recruitment banner.



Figure 13.2: Example of debriefing banner.

Socio-demographics questionnaire

We provide the questions asked to the participants to understand their profiles and backgrounds.

- ¿Cuántos dispositivos usas para conectarte a internet en casa? /How many devices do you use at home to connect to the internet
- ¿Cuántos años tenías la primera vez que usaste internet? / How old were you the first time use used the internet?
- En los últimos 3 meses, ¿cuánto has usado internet? / In the last 3 months, how much have you used the internet?: More than once a week, Once a week, Several times a week, Once a day, More than 2 hours a day, More than 4 hours a day
- ¿En qué año naciste? / In which year you were born?
- ¿Cuál es tu lengua materna? / What is your mother tongue?

- ¿Con qué género te identificas? / With which gender do you identify?
- ¿Cuál es tu profesión? / What is your profession (if you are retired, what was your profession?)
- ¿Dónde naciste? / Where were you born?
- Donde te situarías en la escala socio-económica? Siendo 1 lo más bajo y 10 lo más alto. / Where would you situate yourself in the socio-economic scale? 1 the lowest, 10 the highest.

13.3 Chapter 9

Here, I include the items used from the validation in Spanish [120] of the Internet Skills Scale [444]. I accompany the English translation for transparency purposes.

Items in Spanish	English translation
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Sé cómo abrir archivos que he descargado]	Think about when you use a computer, how much do you agree or disagree with the following statements? [I know how to open files that I have downloaded]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Sé cómo bajar o guardar una foto que haya encontrado online]	Think about when you use a computer, how much do you agree or disagree with the following statements? [I know how to download or save a photo that I found online]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Sé cómo utilizar teclas de acceso rápido o atajos (Ej. Ctrl-C para copiar, Ctrl-V para pegar)]	Think about when you use a computer, how much do you agree or disagree with the following statements? [I know how to use keyboard shortcuts (e.g., Ctrl-C to copy, Ctrl-V to paste)]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Me es sencillo decidir cuáles son las mejores palabras clave a utilizar para hacer una búsqueda online]	Think about when you use a computer, how much do you agree or disagree with the following statements? [It's easy for me to decide which are the best keywords to use for an online search]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Me es sencillo encontrar fácilmente una página web que haya visitado anteriormente]	Think about when you use a computer, how much do you agree or disagree with the following statements? [It's easy for me to find a website that I have previously visited]

Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Disfruto de realizar búsquedas de información en Internet]	Think about when you use a computer, how much do you agree or disagree with the following statements? [I enjoy searching for information on the Internet]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [A veces termino en páginas web sin saber bien cómo he llegado a ellas]	Think about when you use a computer, how much do you agree or disagree with the following statements? [Sometimes I end up on websites without knowing exactly how I got there]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Me resulta sencillo darme cuenta cuán cierta es la información que encuentro en Internet]	Think about when you use a computer, how much do you agree or disagree with the following statements? [It's easy for me to recognize how accurate the information I find on the Internet is]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Sé qué tipo de información debo/no debo compartir online]	Think about when you use a computer, how much do you agree or disagree with the following statements? [I know what type of information I should/should not share online]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Sé cómo abrir archivos que he descargado]	Think about when you use a computer, how much do you agree or disagree with the following statements? [I know how to open files that I have downloaded]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Sé cómo bajar o guardar una foto que haya encontrado online]	Think about when you use a computer, how much do you agree or disagree with the following statements? [I know how to download or save a photo that I found online]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Sé cómo utilizar teclas de acceso rápido o atajos (Ej. Ctrl-C para copiar, Ctrl-V para pegar)]	Think about when you use a computer, how much do you agree or disagree with the following statements? [I know how to use keyboard shortcuts (e.g., Ctrl-C to copy, Ctrl-V to paste)]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Me es sencillo decidir cuáles son las mejores palabras clave a utilizar para hacer una búsqueda online]	Think about when you use a computer, how much do you agree or disagree with the following statements? [It's easy for me to decide which are the best keywords to use for an online search]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Me es sencillo encontrar fácilmente una página web que haya visitado anteriormente]	Think about when you use a computer, how much do you agree or disagree with the following statements? [It's easy for me to find a website that I have previously visited]

Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Disfruto de realizar búsquedas de información en Internet]	Think about when you use a computer, how much do you agree or disagree with the following statements? [I enjoy searching for information on the Internet]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [A veces termino en páginas web sin saber bien cómo he llegado a ellas]	Think about when you use a computer, how much do you agree or disagree with the following statements? [Sometimes I end up on websites without knowing exactly how I got there]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Me resulta sencillo darme cuenta cuán cierta es la información que encuentro en Internet]	Think about when you use a computer, how much do you agree or disagree with the following statements? [It's easy for me to recognize how accurate the information I find on the Internet is]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Sé qué tipo de información debo/no debo compartir online]	Think about when you use a computer, how much do you agree or disagree with the following statements? [I know what type of information I should/should not share online]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Sé cuándo debo/no debo compartir información online]	Think about when you use a computer, how much do you agree or disagree with the following statements? [I know when I should/should not share information online]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Tengo cuidado en hacer que mis comentarios y comportamientos sean los adecuados para cada situación que me encuentro online]	Think about when you use a computer, how much do you agree or disagree with the following statements? [I am careful to ensure that my comments and behaviors are appropriate for each situation I encounter online]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Sé cómo cambiar con quién comparto contenidos online (Ej. amigos, amigos de amigos, o público en general)]	Think about when you use a computer, how much do you agree or disagree with the following statements? [I know how to change who I share online content with (e.g., friends, friends of friends, or the public)]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Sé cómo crear y subir contenido a Internet]	Think about when you use a computer, how much do you agree or disagree with the following statements? [I know how to create and upload content to the Internet]

Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Sé cómo crear algo nuevo a partir de imágenes, canciones y videos que encuentro online]	Think about when you use a computer, how much do you agree or disagree with the following statements? [I know how to create something new from images, songs, and videos I find online]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Sé cómo diseñar una página web]	Think about when you use a computer, how much do you agree or disagree with the following statements? [I know how to design a website]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Sé cómo comprar o pagar por una aplicación desde mi teléfono móvil]	Think about when you use a computer, how much do you agree or disagree with the following statements? [I know how to buy or pay for an app from my mobile phone]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Me preocupo de que mi comportamiento y comentarios en internet sean adecuados a las situaciones en que me encuentro]	Think about when you use a computer, how much do you agree or disagree with the following statements? [I care about ensuring that my behavior and comments on the Internet are appropriate for the situations I encounter]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Sé cómo marcar una página web como favorita]	Think about when you use a computer, how much do you agree or disagree with the following statements? [I know how to bookmark a website]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Sé cómo eliminar amigos de mis listas de contactos]	Think about when you use a computer, how much do you agree or disagree with the following statements? [I know how to remove friends from my contact lists]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Sé cómo abrir una nueva pestaña en un navegador]	Think about when you use a computer, how much do you agree or disagree with the following statements? [I know how to open a new tab in a web browser]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Sé cómo instalar aplicaciones en un teléfono móvil]	Think about when you use a computer, how much do you agree or disagree with the following statements? [I know how to install apps on a mobile phone]
Piensa en cuando utilizas un ordenador, ¿cómo de acuerdo o en desacuerdo estás con las siguientes afirmaciones? [Sé cómo controlar los gastos del uso de Internet de mi teléfono móvil]	Think about when you use a computer, how much do you agree or disagree with the following statements? [I know how to manage the costs of Internet usage on my mobile phone]

