Disentangling Online Manipulation Strategies from the Perspective of Digital Inequalities

LORENA SÁNCHEZ CHAMORRO, University of Luxembourg, Luxembourg

There is a growing concern about the use of manipulative mechanisms in online interfaces and the underlying harm to users' autonomy under the label of 'dark patterns'. The effect of these mechanisms, especially in vulnerable populations, is still under-researched. The present doctoral dissertation aims to understand the conditions under which users are less likely to resist manipulation, following the idea of digital inequalities in manipulative designs. This dissertation expects to contribute in two ways. First, bringing empirical insights about the effectiveness of manipulative mechanisms, especially in low digital skilled populations, and providing enablers that allow them to resist manipulation. Second, providing with guidelines that support practitioners in designing user interfaces for these populations and policymakers in improving the current regulations on manipulative designs.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; • **Social and professional topics** → **User characteristics**; **Computing literacy**.

Additional Key Words and Phrases: Dark patterns, deceptive design, manipulative design, digital inequalities

ACM Reference Format:
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, 7 pages. https://doi.org/10.1145/3544549.3577060

1 RESEARCH MOTIVATION

There is a growing concern about the use of manipulative mechanisms in online interfaces and the underlying harm to users' autonomy, health, privacy, or economy [25]. Users might for instance be influenced to spend more time on social media [28] to buy more goods on e-commerce platforms [34], or to release more personal data to service providers in exchange for free features [8, 16, 24]. Overall, the concern revolves around users being manipulated into decisions they initially would not consider making. This risk is amplified by the intricacies between the online domain and everyday activities, e.g., shopping [29], leisure, social media [14], videogames [11], travelling, and streaming services [6], and the widespread use of data to personalise the design of interfaces [3, 8, 17, 26]. The amplitude of these practices, grouped under the label of 'dark patterns', is catching the attention of policymakers, who recently acknowledged this concern based on empirical academic research and investigations around consumer practices [1, 7, 31, 32]. One main problem is the lack of certainty about the specific features that make these manipulative designs so effective. Not only is there a general dearth of empirical research; the particular problem that manipulative mechanisms impact specific users groups more than others has not been researched, either (e.g., socio-economical background, minorities, children). Not enough attention has been paid so far on these vulnerable groups, which are considered worthy of special protection in regulations.

Manuscript submitted to ACM

The present doctoral dissertation addresses the impact of manipulative interfaces and mechanisms, especially from the perspective of digital inequalities. It aims at gathering empirical insights and deriving interface design guidelines, that can be worthwhile for policymakers to focus their attention on the most crucial issues in the realm of dark patterns.

2 RELATED WORK

Persuasion in the online domain is defined as trying to influence someone's behaviours transparently by using 'rational' arguments and incentives that the persuadee can resist [4, 5, 36]. In opposition to rational persuasion, manipulation refers to subverting the manipulee's vulnerabilities in a hidden way [35, 36]. To resist this influence, prior work has shown that it is necessary to perceive the try of subversion and have agency over it [2, 3, 23, 27]. Therefore, it is first necessary to understand the users' vulnerabilities towards user interfaces to analyse if they are manipulated or not and, second, test their capability to resist it. Users can be vulnerable online, or less likely to resist manipulation, for different conditions -both personal and external-, and the combination of vulnerabilities makes some users more vulnerable than others [21]. Ecological theories explain, for instance, how users perceive online interfaces differently, building on meso-conditions [18], like education levels, environment, and socio-economic conditions, that can contribute to this position of vulnerability.

In the realm of 'dark patterns', policymakers, like the European Commission or the Federal Trade Commission [7], have attempted to narrow the concept down. Manipulative and deceptive patterns are defined as "design practices that impair autonomy", and scholars have developed attributes and taxonomies [9, 15, 24]. Empirical studies have only provided an overview of how some designs can change behaviours. Restricting options in interfaces [30], hiding information, pre-selecting choices or presenting confusing information, or even combining these elements, was shown to be effective at influencing users [20]. Other strategies however (e.g., using emotions, social proof, or scarcity cues [16, 20, 22, 30]) need further analysis to investigate their effects on users. Scholars have demonstrated that users can recognise these practices, but their potential influence or the conditions under which users can resist them are still unclear. Moreover, although the literature on awareness and detection of dark patterns is increasing, the effects these mechanisms have on vulnerable populations -children, the elderly, lower digitally skilled users, and minorities- are under-explored. Bongard-Blanchy et al. [2] and Luguri and Strahilevitz [20] have found manipulative design to be more impactful on users with a lower education level, hinting at a problem that is urgent to address.

Among the potential conditions of vulnerability, it is necessary to analyse the role of these meso-conditions, like socio-digital inequalities. Socio-digital inequalities are *"systematic differences between individuals from different backgrounds in the* ***opportunities and abilities to*** *translate digital engagement into benefits and* ***avoid the harm*** *that might result from engagement with ICTS."* [19](p. 44). In my research, I will consider the digital skills divide to evaluate the differences in the outcomes of users' online engagement, where the outcomes are the differences in resisting manipulation. I will rely on a measure of digital skills as a proxy for digital inequalities and evaluate its role in being manipulated online. Digital skills are well-documented to be impacted by educational levels and economic situation [10, 12, 19, 37–40]. The main hypothesis is that a low level of digital skills may render persuasive elements manipulative.

3 RESEARCH OBJECTIVES

The present research aims to understand the conditions under which users are less likely to resist manipulation, following the idea of digital inequalities in manipulative designs. On the one hand, I aim to explore the design mechanisms of manipulation as an external condition. On the other hand, I aim to investigate the role of digital inequalities in the effects that manipulative designs have on users. I will thus bridge the digital inequalities literature with the HCI domain,

trying to understand under which conditions i.e., both external -manipulative designs- and meso -socio-economic conditions- we can consider users vulnerable online and thus less likely to resist manipulation. Knowing the conditions and perceptions of manipulation in this specific population will allow to derive design recommendations for the development of user interfaces.

This dissertation has the following objectives. First, analysing the persuasive design mechanisms that can become manipulative and the actual industrial practices around them (Study 1). Second, understanding how these manipulative design elements are perceived by users with lower digital skills and how they resist them (Study 2). Third, disentangling the factors that can contribute to the resistance to manipulative elements in an experimental setting, analysing a potential divide in the results (Study 3).

Building on the above, I am addressing the following research questions:

- (RQ1) Under which conditions can persuasive mechanisms in user interfaces become manipulative?
- (RQ2) What is the relationship between digital inequalities and resistance to manipulation?

I will explore three dimensions of the interaction between humans and user interfaces: (1) the design mechanisms adopted by practitioners, (2) how users perceive and are impacted by manipulative interfaces, and (3) the outcomes' divides between digitally low vs. high-skilled users.

Through this work, I intend to make the following contributions to the HCI community:

- Providing insights into what design elements, mechanisms and strategies in user interfaces can be problematic in online manipulation. These can take the form of guidelines and heuristics for practitioners, scholars and policymakers that allow them to self-assess when designing for influencing behaviours.
- Providing empirical insights into how underrepresented populations in HCI, e.g., low digital skilled users, use and perceive manipulative technologies. Identifying the enablers that allow resisting manipulation can support ethical practices in design.
- Bridging digital inequalities literature with HCI. By demonstrating a divide in the interaction with manipulative technologies and the causes of this divide, we will provide novel insights into the circumstances under which users are vulnerable online. These findings will serve as the ground for policymakers to improve current regulations on manipulative design, consumer protection and dark patterns.

3.1 Study 1: Manipulation in Practice: What are the Mechanisms of Manipulative Interfaces?

Given the problem of defining dark patterns and what can be considered problematic for users, I decided to disentangle the differences between persuasive and manipulative mechanisms in UX design practices. Persuasive design is inherent to interaction design, but some mechanisms might become manipulative if they subvert vulnerabilities in a hidden way. This distinction was not easy to distinguish in user interfaces, and is sometimes hidden or unknown in common UX practices by designers. Therefore, I conducted a study (study 1) to determine problematic persuasive design mechanisms that might become manipulative UX practices in a simulated scenario where designers are asked to influence users.

I looked at the UX design practices to engage with their mechanisms of influencing users, and put them together with theories of manipulation. It is a qualitative study (n=22) aimed at answering the following research questions:

- (RQ1) How do designers understand different types of influence in design?
- (RQ2) How do designers implement influence mechanisms in design?
- (RQ3) What contextual factors inform designers' decisions when they are asked to influence online?

We elicited initial heuristics to prevent manipulation, informed by the designers' approaches to influence and the literature review. The Heuristics, therefore, represent design mechanisms that, in combination, can become manipulation. For instance, a non-transparent interface, with a lot of frictional elements that are not supporting users' goals, can be manipulative. In contrast, only a frictional element per se does not necessarily imply a manipulative element.

3.2 Study 2 (envisioned): What Does 'Resistible' Mean for Vulnerable Populations? Perceptions of Manipulation Resistance in Low Digital Skilled Populations

The objective is to understand how low digitally skilled users perceive manipulative techniques, including contextual factors, and if they perceive that manipulative techniques impact their decisions online. It is expected to be a qualitative exploratory study. I will therefore build a qualitative model that gives us an idea of how the design mechanism affects context understanding and resistibility of the incentives as the basis for online manipulation. We aim to make an exploratory analysis of the strategies that might be particularly problematic for specific subpopulations.

- (RQ1) How do low digitally skilled users perceive the use of incentives when manipulated online?
- (RQ2) What are the perceived impacts in their decisions for low digitally skilled users when they are manipulated online?

Study design. This is an exploratory qualitative analysis inspired from [13][33]. Participants will be shown scenarios in which they are presented with familiar situations with manipulative designs -e.g. flight booking, cookie banners, or e-commerce websites. I will conduct a semi-structured interview to explore their perceptions about: (i) how they react against those practices, (ii) what they understand of the mechanism of manipulation, (iii) how they protect themselves from manipulation and (iv) the extent to which they perceive changes in their behaviours caused by those mechanisms. To know the participant's level of digital skills, they will be asked some questions in the screening questionnaire.

3.3 Study 3 (envisioned): What are the Impacts of Digital Skills on Resistance to Manipulation?

The objective of this study is to explore the association between digital inequalities and resistance to manipulation online. Given the strong relationship between socioeconomic inequalities and the digital divides -access, skills and outcomes- that have been widely explored in the literature [12, 19, 37, 39, 40], I will use digital skills as a proxy for digital inequalities. In this regard, the divide in outcomes -or what users can obtain from using technology and being included in society- depends on their online abilities or digital skills. However, there is a strong relationship between how users acquire digital skills and their socio-economic status. In this regard, I will answer the following research questions:

- (RQ1) What is the relationship between digital skills and resistance to manipulation?
- (RQ2) What are the conditions under which users are more likely to resist manipulation?

Therefore, I will explore the relationship between digital skills and resistance to manipulation in two ways. First, I will analyse the existence of the divide and then the different explainable personal factors that will lead to that divide in resisting manipulation: economic background, education or family education.

Study design. Building on my exploration of the qualitative model that explains the idea of resistance manipulation in low digitally skilled users (See Study 2), I will conduct a large-scale quantitative study. Participants will be presented with a scenario where they will have to spend money on a Fast-Fashion or an NGO website. They will be told that after the task, they will keep the money they do not spend on the product. For example, if the compensation is 20 euros and

they pay for a 5 euros product, they will keep the product and only 15 euros instead 20. This is an important aspect to guarantee ecological validity in the scenario. I expect to control the scenario with different incentives according to the qualitative model. I will analyse the differences in behaviour given the level of digital skills, controlling for different socio-economic factors, through difference-in-difference quasi-experimental analysis.

4 NEXT STEPS AND CHALLENGES

The main challenge in finding empirical quantitative evidence to disentangle factors that affect behaviour relates to the validity of the experiments in this specific topic. There are two types of challenges in this regard. The first one is determining the specific design features that work for manipulation. In the online domain, manipulation becomes very context dependent; therefore, it is difficult to attribute in both experimental settings and natural context -e.g. data collection with tracking systems and identifiers online- what are the design elements that affect the user's behaviour. Therefore, I am looking at manipulation mechanisms and strategies to change behaviour. Nonetheless, this challenges operationalising 'resisting the mechanism of influence'.

Setting experimental designs in the online domain with high validity remains problematic and complex, given the difficulty of studying participants in a natural context. Some considerations show a challenge that comes associated with studying online manipulation, given deception becomes necessary in the study design. Therefore, deceiving participants with a high validity in an experimental setting becomes very difficult. Artificial settings present problems given the number of variables that are outside control in a natural context, like the actual willingness or need to acquire a product -in e-commerce settings-; the need for information -in newspaper and news consumption settings- or the actual privacy trade-offs and risks that users make when they have to yield their data online -in the case of cookie banners experiments. When participants are in an artificial setting, they are less likely to consider trade-offs, risks and decisions in the same way they would do in a natural situation. This is one of the common limitations that are presented in the empirical studies in the domain, sometimes leading to limited or inconclusive results. These methodological challenges become specially problematic in the context of low-digital skills users.

REFERENCES

[1] European Data Protection Board. 2022. Guidelines 3/2022 on Dark patterns in Social Media Platform Interfaces: How to Recognise and Avoid Them.

[2] Kerstin Bongard-Blanchy, Arianna Rossi, Salvador Rivas, Sophie Doublet, Vincent Koenig, and Gabriele Lenzini. 2021. I am Definitely Manipulated, Even When I am Aware of it. It s Ridiculous! – Dark Patterns from the End-User Perspective, In ACM DIS Conference on Designing Interactive Systems, June 28– July 2, 2021, Virtual event, USA. *arXiv:2104.12653 [cs]*, 22. arXiv: 2104.12653.

[3] Ida Borberg, Rene Hougaard, Willard Rafnsson, and Oksana Kulyk. 2022. "So I Sold My Soul": Effects of Dark Patterns in Cookie Notices on End-User Behavior and Perceptions. In *Workshop on Usable Security and Privacy (USEC) 2022 3 March 2022, San Diego, California, USA*. 11.

[4] Marietjie Botes. 2022. Autonomy and the social dilemma of online manipulative behavior. *AI and Ethics* (Apr 2022). https://doi.org/10.1007/s43681-022-00157-5

[5] Lorena Sanchez Chamorro, Kerstin Bongard-Blanchy, and Vincent Koenig. 2022. Justice in interaction design: preventing manipulation in interfaces. http://arxiv.org/abs/2204.06821 arXiv:2204.06821 [cs].

[6] Akash Chaudhary, Jaivrat Saroha, Kyzyl Monteiro, Angus G. Forbes, and Aman Parnami. 2022. "Are You Still Watching?": Exploring Unintended User Behaviors and Dark Patterns on Video Streaming Platforms. In *Designing Interactive Systems Conference*. ACM, Virtual Event Australia, 776–791. https://doi.org/10.1145/3532106.3533562

[7] Federal Trade Commission. 2022. Bringing Dark Patterns to Light. , 48 pages.

[8] Lorrie Faith Cranor. 2022. Cookie monster. *Commun. ACM* 65, 7 (Jul 2022), 30–32. https://doi.org/10.1145/3538639

[9] Andrea Curley, Dympna O'Sullivan, Damian Gordon, Brendan Tierney, and Ioannis Stavrakakis. 2021. The Design of a Framework for the Detection of Web-Based Dark Patterns. In *ICDS 2021: The 15th International Conference on Digital Society*. Nice, France (Online), 8.

[10] A. J. A. M. van Deursen and J. A. G. M. van Dijk. 2010. Measuring Internet Skills. *International Journal of Human-Computer Interaction* 26, 10 (Sep 2010), 891–916. https://doi.org/10.1080/10447318.2010.496338

[11] Bruno Dupont and Steven Malliet. 2021. Contextualizing Dark Patterns with the Ludeme Theory: A New Path for Digital Game Literacy? *Acta Ludologica* 4 (2021), 20.

[12] Rebecca Eynon and Anne Geniets. 2016. The digital skills paradox: how do digitally excluded youth develop skills to use the internet? *Learning, Media and Technology* 41, 3 (Jul 2016), 463–479. https://doi.org/10.1080/17439884.2014.1002845

[13] Alisa Frik, Julia Bernd, Noura Alomar, and Serge Egelman. 2020. A Qualitative Model of Older Adults' Contextual Decision-Making About Information Sharing. In *20th Annual Workshop on the Economics of Information Security (WEIS 2020), Brussels, Belgium, December 14–15*. 62.

[14] Liza Gak, Seyi Olojo, and Niloufar Salehi. 2021. The Distressing Ads That Persist: Uncovering The Persuasive Logics and Emotional Harms of User Targeted Diet Ads. Online Virtual Conference (originally Yokohama, Japan), 4.

[15] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. The Dark (Patterns) Side of UX Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–14. https://doi.org/10.1145/3173574.3174108

[16] Paul Graßl, Hanna Schraffenberger, Frederik Zuiderveen Borgesius, and Moniek Buijzen. 2021. Dark and bright patterns in cookie consent requests. 3, 1 (2021), 38.

[17] Frode Guribye, Oda Elise Nordberg, Are Nyhammer, Marija Slavkovik, and Htut Soe. 2021. Dark patterns in cookie consent notices: new definitions and mitigation strategies. (2021), 5.

[18] Ellen Johanna Helsper. 2016. *Survey on the use of information and communication technologies in Brazilian households: ICT households 2015*. São Paulo, Brasil. 175–185 pages.

[19] Ellen J. Helsper. 2021. *The Digital Disconnect. The social causes and consequences of digital inequalities*. SAGE Publications Ltd.

[20] Jamie Luguri and Lior Jacob Strahilevitz. 2021. Shining a Light on Dark Patterns. *Journal of Legal Analysis* 13, 1 (Mar 2021), 43–109. https://doi.org/10.1093/jla/laaa006

[21] Florencia Luna. 2019. Identifying and evaluating layers of vulnerability – a way forward. *Developing World Bioethics* 19, 2 (Jun 2019), 86–95. https://doi.org/10.1111/dewb.12206

[22] Francisco Lupiáñez Villanueva, Alba Boluda, Francesco Bogliacino, and Giovanni Liva. 2022. Behavioural study on unfair commercial practices in the digital environment: dark patterns and manipulative personalisation. , 303 pages.

[23] Maximilian Maier and Rikard Harr. 2020. Dark Design Patterns: An End-User Perspective. *Human Technology* 16, 2 (Aug 2020), 170–199. https://doi.org/10.17011/ht/urn.202008245641

[24] Arunesh Mathur, Gunes Acar, Michael J. Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov 2019), 1–32. https://doi.org/10.1145/3359183 arXiv: 1907.07032.

[25] Arunesh Mathur, Jonathan Mayer, and Mihir Kshirsagar. 2021. What Makes a Dark Pattern... Dark? Design Attributes, Normative Considerations, and Measurement Methods. *arXiv:2101.04843 [cs]* (Jan 2021). https://doi.org/10.1145/3411764.3445610 arXiv: 2101.04843.

[26] Célestin Matte, Nataliia Bielova, and Cristiana Santos. 2020. Do Cookie Banners Respect my Choice? Measuring Legal Compliance of Banners from IAB Europe's Transparency and Consent Framework. http://arxiv.org/abs/1911.09964 arXiv: 1911.09964.

[27] Christian Meske and Ireti Amojo. 2020. Ethical Guidelines for the Construction of Digital Nudges. In *53rd Hawaii International Conference on System Sciences*. 10.

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

[29] Carol Moser, Sarita Y. Schoenebeck, and Paul Resnick. 2019. Impulse Buying: Design Practices and Consumer Needs. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–15. https://doi.org/10.1145/3290605.3300472

[30] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. 2020. Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence. In *CHI Conference on Human Factors in Computing Systems*. 13. https://doi.org/10.1145/3313831.3376321 arXiv: 2001.02479.

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

[32] Authority on Consumers and Markets (ACM). 2020. Protection of the Online Consumer Boundaries of Online Persuasion.

[33] A.J. Scheerder, A.J.A.M. van Deursen, and J.A.G.M van Dijk. 2020. Taking advantage of the Internet: A qualitative analysis to explain why educational background is decisive in gaining positive outcomes. *Poetics* 80 (Jun 2020), 101426. https://doi.org/10.1016/j.poetic.2019.101426

[34] Ray Sin, Ted Harris, Simon Nilsson, and Talia Beck. 2022. Dark patterns in online shopping: do they work and can nudges help mitigate impulse buying? *Behavioural Public Policy* (May 2022), 1–27. https://doi.org/10.1017/bpp.2022.11

[35] Daniel Susser and Vincent Grimaldi. 2021. Measuring Automated Influence: Between Empirical Evidence and Ethical Values. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Virtual Event USA, 242–253. https://doi.org/10.1145/3461702.3462532

[36] Daniel Susser, Beate Roessler, and Helen F. Nissenbaum. 2019. Online Manipulation: Hidden Influences in a Digital World. *Georgetown Law Technology Review* 4, 1 (2019). https://doi.org/10.2139/ssrn.3306006

[37] Alexander van Deursen and Jan van Dijk. 2011. Internet skills and the digital divide. *New Media & Society* 13, 6 (Sep 2011), 893–911. https://doi.org/10.1177/1461444810386774

[38] Alexander J.A.M. van Deursen, Ellen J. Helsper, and Rebecca Eynon. 2016. Development and validation of the Internet Skills Scale (ISS). *Information, Communication & Society* 19, 6 (Jun 2016), 804–823. https://doi.org/10.1080/1369118X.2015.1078834

[39] Jan van Dijk and Kenneth Hacker. 2003. The Digital Divide as a Complex and Dynamic Phenomenon. *The Information Society* 19, 4 (Sep 2003), 315–326. https://doi.org/10.1080/01972240309487

[40] Jan van Dijk and Alexander J. A. M. van Deursen. 2014. *Digital Skills.* Palgrave Macmillan US, New York. https://doi.org/10.1057/9781137437037