Rules of Engagement: Levelling up to Combat Unethical CUI Design

THOMAS MILDNER, University of Bremen, Germany

GIAN-LUCA SAVINO, University of St.Gallen, Switzerland

PHILIP DOYLE, University College Dublin, Ireland

While a central goal of HCI has always been to create and develop interfaces that are easy to use, a deeper focus has been set more recently on designing interfaces more ethically. However, the exact meaning and measurement of ethical design has yet to be established both within the CUI community and among HCI researchers more broadly. In this position paper we propose a simplified methodology to assess interfaces based on five dimensions taken from prior research on so-called dark patterns. As a result, our approach offers a numeric score to its users representing the manipulative nature of evaluated interfaces. It is hoped that the approach - which draws a distinction between persuasion and manipulative design, and focuses on how the latter functions rather than how it manifests - will provide a viable way for quantifying instances of unethical interface design that will prove useful to researchers, regulators and potentially even users.

Additional Key Words and Phrases: ethical design, dark patterns, conversational user interfaces

1 INTRODUCTION

Although ethical interface design has been a concern in HCI for some time, researchers are only beginning to formulate ways to understand, define and measure occurrences of unethical interface design. In terms of attempts to build knowledge around this issue, most progress can be found in literature on dark patterns, which began just over a decade ago. In 2010, Brignull introduced the term of dark patterns as *"tricks used in websites and apps that make you do things that you didn't mean to"* [3]. Throughout the past decade, many examples of dark patterns have been identified, generating a rich taxonomy of specific dark pattern types and dark pattern strategies [2, 3, 7, 10, 11, 13, 19, 31]. While this is a positive move in terms of better understanding unethical interface design practises, the body of work suffers somewhat from a lack of consensus, continuity, and clear definition about what exactly is being examined across various different studies.

This lack of consensus and continuity in how concepts are defined also echoes a general problem in human-machine dialogue (HMD) research, which has been highlighted in a number of recent reviews of research in the field [4, 5, 28, 29]. The aim of this provocation paper is to draw attention to the potential for learning from the problems within these bodies of work and to further encourage efforts to bring clarity to how we define and measure examples of unethical design in CUIs. As a research community, we might then be able to establish more effective ways of identifying and

measuring unethical design practises in the context of CUI development before they become as common and problematic as they already in graphical interface interaction.

2 FUNDAMENTALS: PERSUASION & MANIPULATION

To a certain degree persuasion is a fundamental feature of design. Ideally, the aim is to design objects that signify their potential uses and constraints so people can interact with them as intuitively as possible. That is, we look to create objects that gently nudge people toward using them in a certain way. This is a common general understanding of design in HCI research, and echoes widely known basic principles for ensuring alignment between a user's mental model and a system forwarded by Norman in 'The Design of Everyday Things' [22]. Yet, in recent years the notion of persuasive design has begun to acquire negative connotations, largely due to a rise in design techniques aimed at governing and exploiting peoples' decision-making. Thaler and Sunstein provide an example of this type of design technique when introducing the term of Nudges [17] to describe interventions that alternate peoples' decision making process in a predictable way, allowing design to be used to navigate a users' focus into a predefined direction or goal. However, approaches of this nature, that take advantage of our cognitive biases, can be used equally efficiently for benevolent or malevolent ends; as is evidenced by the growth in research on dark patterns for both GUIs and CUIs, and across numerous domains.

This raises questions about how we define persuasive design, and indeed, how we draw distinctions between designs regarded as persuasive vs. manipulative. Nudges have certainly been used in beneficial ways, including improving peoples' eating [27], healthcare [32], fitness [21], sleep [15] and relaxation practices [30]. It would also be inaccurate to assume people are naive to how these systems work, with the implicit nature of persuasive cognitive approaches being both part of the appeal and part of the reason they are effective [18]. However, they also share fundamental similarities with designs used to encourage users to make potentially harmful decisions; so-called dark patterns. They both disarm a person of their autonomy, albeit temporarily, by encouraging them to make choices they might not have made otherwise. Therein lies our ethical obligation and challenges as HCI researchers: to ensure we combat unethical acts of manipulative interface design, whilst ensuring we continue to deliver persuasive interface designs that improve the lives of people who use them. We argue that establishing clear conceptual definitions is the first step toward consistent identification and measurement of unethical interface design in CUIs.

Based on established psychological definitions, we suggest drawing a distinction between persuasive design and manipulative design. Here, persuasion is defined as, *"an active attempt by one person to change another person's attitudes, beliefs, or emotions associated with some issue, person, concept, or object"* [23]. To bring this into context with HCI research, we suggest a slight adaptation, defining persuasive design as, *'an active attempt to influence a person's behaviours, attitudes, beliefs, or emotions associated through interface design'*. On the other hand, manipulation is described in psychology as, *'behaviour designed to exploit, control or otherwise influence others to one's own advantage'* [24]; or in the context of HCI, *'designs aimed at exploiting, controlling or otherwise influencing users to one's own advantage'*. Our challenge in bringing clarity to this distinction lies in developing ways to identify and measure occurrences of unethical manipulative design, which will aid us in defending the benefits of using psychological knowledge to improve people's lives through technology also.

3 UNDERSTANDING AND MEASURING UNETHICAL DESIGN: LESSONS FROM DARK PATTERNS

Research into unethical interface design practises began with examinations of e-commerce websites by Brignull [3]. The work identified twelve specific examples of unethical design aimed at inhibiting people's ability to make informed

choices. These include, *Sneak into Basket*, *Hidden Costs*, and *Price Comparison Prevention*, which all operate on the premise of obscuring information and potentially misleading users into buying unwanted or unnecessarily expensive products. Other dark patterns defined by the author include, *Forced Continuity*, *Privacy Zuckering*, and *Roach Motel*, which are all tactics aimed at forcing people to sign-up, or stay signed up for accounts and services they might not require anymore. These examples are also used by service providers to gain access to private data without fully informing users why it is being collected, who has access to it, or even what it might be used for. Since then, researchers have described and defined a plethora of other examples, with a recent review of the literature from Mathur et al. [20] identifying 62 specific types of dark patterns. Understanding how these existing dark patterns were established and how they are used to facilitate unethical design in graphical user interfaces (GUIs) could allow us to prevent similar developments in CUIs.

While it may seem like unethical GUI design has been a perennial problem for HCI researchers, literature on dark patterns shows that these tactics and the form they take develop and change over time through conscious efforts made by interface design practitioners. Grey et al. [9, 11] identified multiple dark patterns as well as design constraints of practitioners which lead to their creation. In a first study, the authors analysed an image-based corpus to define five types of dark patterns that practitioners engage in when developing manipulative designs [11]. Most of Gray at al.'s dark patterns include descriptions from prior works. For example, the *Obstruction* dark pattern, used to make processes unnecessarily difficult, incorporates Brignull's *Roach Motel*, *Price Comparison Prevention*, and *Intermediate Currency* [3]. In a follow up to this work, Gray et al. also analysed 4775 user-generated posts of the Reddit sub-forum *r/assholedesign* [9]. Analysis resulted in identifying six properties of "asshole designers". The work is particularly useful for understanding the origins of dark patterns and how they emerge from constraints under which practitioners work.

While many of the previously described dark patterns are applicable to different domains, research has also found domain specific examples. For example, Zagal et al. identified seven dark patterns that related to video game mechanics [31]. While certain patterns exploit a game's ecosystem of connected users, such as *Social Pyramid Schemes* and *Impersonation*, others impact game-play experience like *Grinding* and *Playing by Appointment*. Elsewhere, Greenberg et al. [13] consider dark patterns in conjunction with proxemics theory [14]. Identifying nine types of dark pattern in total, the authors discuss interactions with manipulative design in spatial environments. The *Attention Grabber* and *Disguised Data Collection* dark patterns, for instance, could be used in the design of digital billboards that exploit people's proximity and personal data to deliver personalised advertising.

Understanding the creation of dark patterns and analysing their occurrences helps to close the "cultural lag" [25] where the creation of ethical guidelines inevitably lags behind the release of novel dark patterns and even novel technologies. However, recent attempts by the authors to apply the aforementioned corpus of dark patterns in the domain of social media highlight a central problem in this body of work: it is hallmarked by a high degree of overlap between definitions, inconsistent terminology, and descriptions that operate on different levels often without explicit acknowledgement. That is, the difference between specific dark pattern designs and broader dark pattern strategies is not always recognised. Attempts to apply taxonomies of dark patterns also shows that while some established dark patterns types are applicable to other domains, it is also likely that novel CUI specific dark patterns types and strategies will need to be described and defined. Further, while this is particularly problematic from a research perspective, we see an even more urgent need to begin this work so we might also better protect users.

4 WAYS TO COMBAT UNETHICAL INTERFACE DESIGN

With regards to technology, particularly online interfaces, studies have shown that the burden to counteract dark patterns often falls on users [1, 8]. Although regulations, such as the GDPR [6] or the CCPA [16], aim to protect users

Mathur 2019 [19]

Dark Pattern Characteristics

Characteristic	Question
Asymmetric	Does the user interface design impose unequal weights or burdens on the available choices presented to the user in the interface?
Covert	Is the effect of the user interface design choice hidden from the user?
Deceptive	Does the user interface design induce false beliefs either through affirmative mis-statements, misleading statements, or omissions?
Hides Information	Does the user interface obscure or delay the presentation of necessary information to the user?
Restrictive	Does the user interface restrict the set of choices available to users?

Table 1. This table lists the introductory questions Mathur et al. (2019) [19] gave for each dark pattern characteristic.

in online environments, the previously mentioned cultural lag [25] means these efforts are struggling to counter all problematic designs described under the umbrella of manipulative design. Indeed, one could argue they have led to the creation of new dark patterns. The design of cookie consent banners that favour 'accept all' options over offering users greater control, shows how design can be used to easily negate efforts to combat dark patterns, and how current regulatory efforts fail to address the fundamental nature of manipulative design. This has lead to designs exploiting cognitive biases not covered by the GDPR, such as anchoring effects, to steer users into giving consent that they might not have given were they provided with a neutral choice [12, 19].

In January 2022, the European Union proposed a new article 13a, as part of the Digital Service Act [26], to address previous concerns. Offering a generalised definition of problematic design that is closely worded to Mathur et al's. [20] definition of dark patterns, article 13a contains an extendable list of specific interface designs to be regulated. However, we argue that this approach may lead to a similar problem seen in attempts to apply dark pattern taxonomies across different interfaces and domains. That is, the taxonomies quickly become very large, difficult to maintain, and not always appropriate across interfaces types and use cases. Further, creating an ever growing list of examples may also deepen this problem over-complicating crucial efforts to combat unethical design.

5 LEVELLING UP TO UNETHICAL INTERFACE DESIGN

By categorising specific dark pattern types into five broad characteristics, Mathur et al. [19] offer an alternative and potential useful avenue for combating manipulative designs. This higher level categorisation is based on the cognitive biases specific designs are developed to exploit. By relying on more fundamental concepts, which are much less interchangeable than specific interface layouts, the approach offers a way to understand manipulative design that focuses on their impact on users, in a broadly applicable fashion, whilst remaining agnostic to the interface or domain in question.

In an attempt to advance their model and differentiate between manipulative and persuasive designs (i.e, between dark patterns and bright patterns), we developed a technique that allows us to evaluate individual examples based on characteristics that stem from the cognitive biases they target. We therefore consider each of the five characteristic by
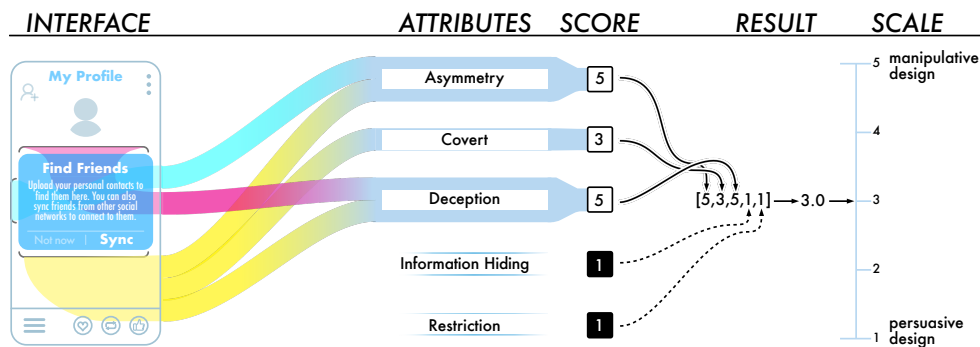
Fig. 1. This figure demonstrates the derivation of a five-dimensional dark pattern vector following Mathur et al.'s five dark pattern characteristics and attributes [19]. Although the example is based on a screenshot of a mobile interface, it could be easily adapted for any conversational user interface.

asking a specific question to assess impact on each of these dimensions, as seen in Table 1. By assigning each a value from 1 (not at all) to 5 (extremely), we are able to evaluate persuasive designs across the five dimensions, indicating the degree to which they might be regarded as persuasive or manipulative. The benefit of this approach is that we gain an overall score that can be used to determine the degree to which a specific design is either persuasive or manipulative, whilst identifying which dimensions of persuasion a manipulative design targets. By providing this clearly defined and measurable conceptualisation of persuasive and manipulative design we might also help regulators combat unethical interface design practices in a consistent and broadly applicable fashion, whilst protecting the benefits of persuasion.

6 CONCLUSION

In this provocation piece, we argue that current attempts to account for manipulative interface design are curtailed by a lack of clarity and continuity around how concepts are defined; hampering efforts to measure and combat their use. We also argue that there are valuable lessons to learn from previous work on GUI dark patterns that might help us head off these problems in the realm of CUIs. Indeed, the approach we suggest may prove useful across multiple domains and interface contexts, and stands to benefit researchers, regulators and potentially users also. Further, by aligning the conceptualisation with how manipulative designs function, rather than how they manifest, we hope to make it much more difficult for designers to circumnavigate. Manipulative design that targets our cognitive biases represents a real and pertinent danger to people, and difficulties faced, even by knowledgeable users and regulators, highlight an urgent need to develop quantifiable approaches that can be easily understood by a range of stakeholders involved in the fight against unethical interface design practices.

REFERENCES

[1] Kerstin Bongard-Blanchy, Arianna Rossi, Salvador Rivas, Sophie Doublet, Vincent Koenig, and Gabriele Lenzini. 2021. "I Am Definitely Manipulated, Even When I Am Aware of It. It's Ridiculous!" - Dark Patterns from the End-User Perspective. In *Designing Interactive Systems Conference 2021* (Virtual Event, USA) *(DIS '21).* Association for Computing Machinery, New York, NY, USA, 763–776. https://doi.org/10.1145/3461778.3462086

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

[3] Harry Brignull, Marc Miquel, Jeremy Rosenberg, and James Offer. 2010. Dark Patterns-User Interfaces Designed to Trick People. http://darkpatterns.org/ visited on 18-02-2022.

[4] Birgit Brüggemeier, Michael Breiter, Miriam Kurz, and Johanna Schiwy. 2020. User Experience of Alexa when controlling music: comparison of face and construct validity of four questionnaires. In *Proceedings of the 2nd Conference on Conversational User Interfaces.* 1–9.

[5] Leigh Clark, Philip Doyle, Diego Garaialde, Emer Gilmartin, Stephan Schlögl, Jens Edlund, Matthew Aylett, João Cabral, Cosmin Munteanu, Justin Edwards, et al. 2019. The state of speech in HCI: Trends, themes and challenges. *Interacting with Computers* 31, 4 (2019), 349–371.

[6] European Commission. 2016. GDPR-16 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-changes_en.pdf

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

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

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

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

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

[12] Colin M. Gray, Cristiana Santos, Nataliia Bielova, Michael Toth, and Damian Clifford. 2021. *Dark Patterns and the Legal Requirements of Consent Banners: An Interaction Criticism Perspective.* Association for Computing Machinery, New York, NY, USA, pp. 1–18. https://doi.org/10.1145/3411764.3445779

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

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

[15] Alexandra Hosszu, Daniel Rosner, and Michael Flaherty. 2019. Sleep Tracking Apps' Design Choices: A Review. In *2019 22nd International Conference on Control Systems and Computer Science (CSCS).* IEEE, 426–431.

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

[17] Thomas C. Leonard. 2008. Richard H. Thaler, Cass R. Sunstein, Nudge: Improving decisions about health, wealth, and happiness. *Constitutional Political Economy* 19, 4 (Dec. 2008), 356–360. https://doi.org/10.1007/s10602-008-9056-2

[18] BGDA Madhusanka and Sureswaran Ramadass. 2021. Implicit intention communication for activities of daily living of elder/disabled people to improve well-being. In *IoT in Healthcare and Ambient Assisted Living.* Springer, 325–342.

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

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

[21] EDIZIONI MINERVA MEDICA. 2018. Fitness mobile apps positively affect attitudes, perceived behavioral control and physical activities. *The Journal of sports medicine and physical fitness* (2018).

[22] Donald A. Norman. 2002. *The design of everyday things.* Basic Books, [New York]. http://www.amazon.de/The-Design-Everyday-Things-Norman/dp/0465067107/ref=wl_it_dp_o_pC_S_nC?ie=UTF8&colid=151193SNGKJT9&coliid=I262V9ZRW8HR2C

[23] APA Dictionary of Psychology. 2022. Manipulation. https://dictionary.apa.org/persuasion visited on 18-02-2022.

[24] APA Dictionary of Psychology. 2022. Manipulation. https://dictionary.apa.org/manipulation visited on 18-02-2022.

[25] W.F. Ogburn. 1922. *Social Change with Respect to Culture and Original Nature.* B.W. Huebsch, Incorporated, University of California.

[26] European Parliament. 2022. Digital Servicess Act***I. European Parliament [A9-0356/2021]. https://www.europarl.europa.eu/doceo/document/TA-9-2022-0014_EN.html

[27] Alessandra Sarcona, Laura Kovacs, Josephine Wright, and Christine Williams. 2017. Differences in eating behavior, physical activity, and health-related lifestyle choices between users and nonusers of mobile health apps. *American Journal of Health Education* 48, 5 (2017), 298–305.

[28] Katie Seaborn and Jacqueline Urakami. 2021. *Measuring Voice UX Quantitatively: A Rapid Review.* Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411763.3451712

[29] Katie Seaborn and Jacqueline Urakami. 2021. Measuring voice UX quantitatively: A rapid review. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–8.

[30] Taelyr Weekly, Nicole Walker, Jill Beck, Sean Akers, and Meaghann Weaver. 2018. A review of apps for calming, relaxation, and mindfulness interventions for pediatric palliative care patients. *Children* 5, 2 (2018), 16.

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

[32] Jing Zhao, Becky Freeman, Mu Li, et al. 2016. Can mobile phone apps influence people's health behavior change? An evidence review. *Journal of medical Internet research* 18, 11 (2016), e5692.