Uncovering Dark Patterns in Persuasive Technology

Agnis Stibe[1], Kate Pangbourne[2], Simon Wells[3], Brian Cugelman[4,5],
and Anne-Kathrine Kjær Christensen[6]

[1] Paris ESLSCA Business School, 75007 Paris, France
`a.stibe@eslsca.fr`
[2] Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, UK
`k.j.pangbourne@leeds.ac.uk`
[3] School of Computing, Edinburgh Napier University, Edinburgh, EH10 5DT, UK
`s.wells@napier.ac.uk`
[4] Statistical Cybermetrics Research Group, University of Wolverhampton, Wolverhampton, UK
[5] AlterSpark, Toronto, ON, Canada
`brian@alterspark.com`
[6] Specifii, Copenhagen, Denmark
`anne@specifii.dk`

Abstract. Dark patterns are interactive design patterns that influence technology users through deception or trickery, and which represent unethical applications of persuasive technology. However, our ability to identify dark patterns is limited, creating a situation where it is difficult to manage abuses of persuasive psychology, because it is difficult to even identify them. Although there are numerous practitioner taxonomies of dark patterns, there is no scientifically-based taxonomy available. This workshop provides an introduction to dark patterns and an overview of the psychological mechanisms that drive them. Through participatory exercises, participants will help to identify the theoretical underpinnings that drive dark patterns, and contribute to the development of a taxonomy of dark patterns, based on consensus within the scientific community. In the workshops, we will form working teams who will review the dark pattern taxonomy, looking for alternative theoretical explanations. Each working team will participate in a group sorting exercise, designed to inform the development of a theoretically-framed taxonomy of dark patterns. All outputs of the workshop will be captured, and used to advance this study towards validation of the taxonomy. After the workshops, the authors of this paper will incorporate all the advancements into the next stage of the research, which will feed into a subsequent paper on a taxonomy of dark patterns, addressing the identified research questions.

Keywords: Dark Patterns, Persuasive Technology, Ethics, Misfires, Persuasive Backfiring, Anti-Patterns, Transforming Sociotech Design.

1   Background

Dark patterns are interactive design patterns that influence technology users through deception, trickery or hostility, that make their lives difficult or contribute a negative

impact, through intended or unintended design practices that represent unethical applications persuasive technology [2]. Examples of dark patterns include getting people to purchase unnecessary insurance, signing up for products without knowing they are on recurring billing, exposing users to content that makes them feel bad about themselves in order to influence their behavior, environmental designs that are effectively 'hostile' to particular groups such as the homeless, cyclists or pedestrians.

Industry professionals have raised public awareness of dark patterns, and have taken steps to identify, collect, and describe dark patterns. Their efforts have helped draw attention to dark patterns [6]. However, there is no comprehensive practitioner taxonomy, they tend to lack a strong theoretical basis, and the taxonomies put forward draw on colloquial terms, rather than frameworks typically employed by behavioral sciences.

2 Relevance to Persuasive Technology

There is a practical need to develop a science-based taxonomy of dark patterns that makes stronger links to behavioral science principles [4]. The existing taxonomies [3] can be greatly improved, through taking a more systematic approach that better identifies the theoretical underpinnings. We extend our definition of persuasive technology beyond the digital to illustrate a long history of darkness in persuasion, our idea is that our taxonomy should be holistic and applicable to different persuasive contexts. As the Internet of Things reaches further into our lives by digitally connecting physical objects and infrastructures, the persuasive effects of smart cities, smart transport and smart homes are likely to be profound.

For this reason, we have been carrying out a systematic review of information about programs, apps, behavioral and environmental designs that may contain dark patterns. This study is aimed at improving our understanding of persuasive technology by better describing what dark patterns are, identifying their theoretical underpinnings, and developing precise definitions that will both describe how to identify them, and how to describe when they are being applied [5]. We are also considering the ethical status of dark patterns in persuasive technology [1].

3 Goals and Research Direction

This workshop will support two goals. First, it provides an opportunity for participants to share knowledge regarding dark patterns, expressed appropriately, for example using visual depictions where necessary, and to discuss the psychological principles that explain their persuasiveness. Second, it will use participatory method, not just to engage participants, but also, to advance this project's goals of developing a taxonomy of dark patterns, that is more rooted in the behavioral science, by making stronger links between theory and practice. Participants will also be equipped to consider whether the use of dark patterns is ever ethically acceptable.

For the workshop, we have prepared a list of research questions to be discussed and further investigated:

- What taxonomy of dark patterns aligns with the behavioral sciences?

- What are the psychological mechanism driving dark patterns?
- How dark patterns differ from anti-patterns (backfires, misfires, see **Fig. 1**)?
- Why do people believe a design pattern constitutes a dark pattern?
- How to define the boundaries from persuasive to manipulative to coercive?
- Are there shades of darkness? How do we detect instances of dark patterns?
- When does a dark pattern and a psychological backfire overlap?
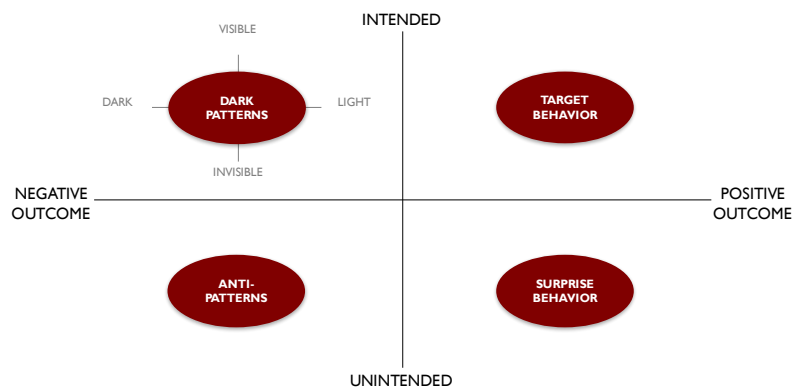- What are the emotional impacts of dark patterns?

Fig. 1. Intention-Outcome matrix adopted from Stibe and Cugelman [4].

4   Examples

We have gathered real-life examples from persuasive and environmental technology that we or our informants regard as dark. A small sample is shown in Table 1 below. Arguably, whilst users still have choices, by linking these examples to behavioral theory we intend to show that unconscious processes are triggered first and more strongly than the conscious processes required to make informed choices. The balance of knowledge is asymmetric in favor of giving the provider more power.

Table 1. Examples of potential dark patterns.

Candidate	Rationale
Cycling infrastructure: poor designs.	Cyclists interpret poor design as deliberate hostility.
Mobile phone data contracts.	Once data limit is reached, some providers continue to provide data but at a higher cost, busting user budgets.
Smartphone setting defaults.	Autoplay default encourages data burning. Updates may replace user settings with default.
Social media ad-targeting.	Political profiling using personal user data then targeting users with paid partisan ads.
Social credit score system.	Social credit scoring utilizing data on users and contacts.

5   Structure and Outcomes

The workshop will start with an introduction followed by presentations on cognate topics, designed to provide background for the interactive exercise, overall:

- Ethics of persuasive technology,
- Resistance against manipulative behavior (trust in source, oxytocin and emotion),
- Grey areas in our understanding of dark patterns,
- Taxonomy of dark patterns (overview sheet and cards),
- Interactive classification exercise (like delphi or Q-method),
- Review of the data, discuss, and clean up the taxonomy based on group consensus.

Prior to the workshop, our team will have completed a systematic review of dark patterns, collecting examples in narrative and visual format, drawing on academic and practitioner sources. These will be aggregated into a list, that includes both simple examples with clear links to persuasion theory, and those that are more complex and not as easy to identify. Using expert review and consensus building, our team will develop the first taxonomy, that links dark patterns to principles routinely used in behavioral science, based on a grounded theory methodology, finalized through an expert review and consensual agreement. The final output will be a list of dark patterns, linked to theory and with an example of each pattern. These will be printed onto cards, so that each table in the workshop will have a list of dark patterns.

In the workshops, we will form working teams who will review the dark pattern taxonomy, looking for alternative theoretical explanations. Each working team will participate in a group sorting exercise, designed to inform the development of a theoretically-framed taxonomy of dark patterns. All outputs of the workshop will be captured, and used to advance this study towards validation of the taxonomy. After the workshops, the authors of this workshop will incorporate all the advancements into the next stage of the research, which will feed into a subsequent scientific publication on a taxonomy of dark patterns, addressing the identified research questions.

References

1. Berdichevsky, D., Neuenschwander, E.: Toward an Ethics of Persuasive Technology, Communications of the ACM, 42 (5), 51-58 (1999).
2. Fogg, B.J.: Persuasive Technology: Using Computers to Change What We Think and Do. San Francisco: Morgan Kaufmann (2003).
3. Spahn, A.: And Lead Us (not) into Persuasion…? Persuasive Technology and the Ethics of Communication. Science and Engineering Ethics, 18(4), 633-650 (2012).
4. Stibe, A. and Cugelman, B.: Persuasive Backfiring: When Behavior Change Interventions Trigger Unintended Negative Outcomes. In International Conference on Persuasive Technology, pp. 65-77. Springer (2016).
5. Verbeek, P.P.: Persuasive Technology and Moral Responsibility Toward an Ethical Framework for Persuasive Technologies. Persuasive, 6, 1-15 (2006).
6. Zagal, J. P., Björk, S., Lewis, C.: Dark Patterns in the Design of Games. In Foundations of Digital Games (2013).