When Persuasive Technology Gets Dark?

Tobias Nyström[1] and Agnis Stibe[2,3]

[1] Uppsala University, Sweden
[2] Métis Lab, EM Normandie Business School, Paris, France
[3] INTERACT Research Unit, University of Oulu, Finland

Abstract. Influencing systems and persuasive technology (PT) should give their users a positive experience. While that sounds attractive and many rush implementing novel ideas things such as gamification, a serious professional and scientifically rich discussion is needed to portray a holistic picture on technology influence. Relatively little research has been aimed at exploring the negative aspects, outcomes, and side effects of PT. Therefore this research aims at addressing this gap by reviewing the existing knowledge on dark patterns, demonstrating how intended Pt designs can be critically examined, introducing the Visibility-Darkness matrix to categorize and locate dark patterns, and proposing a Framework for Evaluating the Darkness of Persuasive Technology (FEDPT). The framework is instrumental for designers and developers of influential technology, as it clarifies an area where their products and services can have a negative impact on well-being, in other words, can become harmful to the users.

Keywords: Dark Patterns · Design · Evaluation · Framework · Negative · Persuasive Technology · Visibility-Darkness Matrix

1 Introduction

Like most technological advancements, an introduction and use of persuasive technology (PT) can have both beneficial and harmful effects on the users. Game experience has recently gained rapid popularity as an enabler of persuasion, as it drives the engagement by using game elements. Oftentimes, game experience is the catalyst for increasing the efficiency of the designed and intended persuasion. However, when it comes to real-life implementations, all technologies can potentially be used for good or bad. Moreover, the study of unintended negative consequences of behavioral interventions is growing and becoming an important research area [16, 17, 38, 48].

In persuasion and computer game research, the harmful effects on people are often labeled as "the dark side" [4, 12, 23, 28, 31], and the issues of ethics are rarely explored [30]. Therefore, it is important to thoroughly study both direct and indirect effects of PT in the context of darkness.

Previous research on negative user effects and human behavior (e.g. the dark side and persuasive backfiring) calls for further exploration [4, 48]. This research aims at advancing this essential but relatively uncharted area of PT. Thus, the

"The final authenticated version is available online at
https://doi.org/10.1007/978-3-030-63396-7_22"

research question for this work is: *When and how can PT get dark?* To address this question, the Visibility-Darkness matrix is proposed and used to identify PT that can be classified as manipulative or designed with bad intentions.

The paper is structured in six sections. The background in section 2 presents the concept of PT, its negative sides, and dark patterns. In section 3 an outline of search results is being mapped into the Visibility-Darkness matrix. A validation of the framework by using use case is shown in section 4 followed by a discussion in section 5. Finally in section 6 conclusions and future research paths are given.

2 Background

2.1 Persuasive technology

The design and use of PT for transforming human behavior can be done in various forms, for example, to help smoking cessation, exercising frequently, driving less and biking more [26]. Fogg defined persuasion in the context of persuasive computers as "an attempt to shape, reinforce, or change behaviors, feelings, or thoughts about an issue, object, or action" [18]. Later in the context of using computers as persuasive technologies (captology) Fogg [19] define persuasion "as an attempt to change attitudes or behaviors or both (without using coercion or deception)", the intended change and planned persuasive effects are central in captology. As coercion is an antonym of persuasion, any technology using force should be labeled as coercive technology. Adding deception to the definition as does not make the PT unproblematic, as ethical issues often emerge in the design phase [40]. Important to note is that both coercion and deception can be a subjective experience for an individual. One popular example of a PT is the use of gamification to change behavior (e.g. persuade towards sustainability see [39]). Commonly gamification is defined as the use of game elements in a non-game context [13] or as a process of providing affordances for gameful experiences which support the customers' overall value creation [27]. Gamification is usually rich with applying points, badges, leaderboards and often includes progress bars, quests, avatars, and performance graphs.

2.2 Possible pitfalls when designing persuasive technology

A literature search using Scopus and Web of Science (2018-10-05) with the keywords ("persuasive technolog*" OR "persuasive system*") AND (dark* OR backfir* OR negative OR ethics OR manipulation OR exploitation) was conducted. It was done to identify prior findings about negative outcomes (and synonyms like backfire, backfiring, darkness, dark side, ethics, exploitation, and manipulation) of persuasive technolog(y/ies)/system(s). Scopus returned 90 papers and Web of Science returned 45 papers and the total number of unique papers was 98. The key inclusion criteria was defined as: peer-reviewed research that address negative effect(s) of PT on individual users. In the first round after reading the abstracts, 32 papers were chosen as candidates, thus further read in

Table 1: Papers exploring pitfalls of persuasive technology

Id	Theme	Id	Theme
[5]	Awareness of unintended outcomes	[6]	Privacy and designer responsibility
[22]	Responsiblity and ethical consideration	[24] [25]	Morally acceptable
[29]	Ethical consideration of adaptable PT	[30]	Ethical framework gamification
[35]	Applicability of discourse ethics on PT	[36]	Applying discourse ethics on PT
[42]	Awareness - lack of understanding and commitment	[43]	Ethical acceptability of PT
[45]	PT design concerns privacy, autonomy, and coercion	[47]	Design guidelines by using discourse ethics
[48]	Awareness and a taxonomy for PT	[49]	Investigate the moral acceptability of machine persuasion
[53]	Autonomy and volunteerism to PT	[55]	Autonomy and volunteerism to PT
[56]	Critical design questions to assess value, action, and goal		

detail. In the second round, the whole paper was read and 18 papers fulfilled the key inclusion criteria, so had relevance for this research, see Table 1.

By reading the papers a few themes where discovered. A number of papers [5, 6, 22, 30, 45, 55] discuss the ethics and responsibility of PT from different viewpoints. Unintended outcomes of PT are discussed e.g. compusuasion is Atkinson's [5] term for unintended behavior change. Fogg's [19] focus on the intended outcome of PT and omission of unintended outcome of PT is problematic since the latter could have a large impact. According to Atkinson [5], the designer of PT should take responsibility for unintended, unforeseen, and unpredicted outcome, although they could not categorically be seen as belonging to persuasion.

Berdichevsky and Neuenschwander [6] explore the ethics of PT by suggesting a systematic approach and develop a framework for evaluating the ethics of the interaction of persuader, PT and the persuaded. They also display a flowchart that shows how ethical responsibility is connected to predictable/unpredictable intent and intended/unintended outcomes. As a summary, their work outlines ten principles for ethical design of PT. Gram-Hansen [22] also explores ethics of PT, especially the impact of ubiquitous technology, as it probably is the most efficient way to change user's behavior. The problem with ubiquitous technology is that it could change human behavior without proper disclosure. This calls for an ethical consideration during the whole design process and evaluation of both the original intention and the practical application. Kim and Werbach [30] identified several ethical issues that needs to be addressed. The issues are framed into four categories: exploitation, manipulation, harms, and (detrimental to) character. These four issues could be the base to formulate a framework for evaluation. Reitberg et al. [45] looked at PT design concerns related to ethics. The TV Com-

panion application (aimed to persuade users towards healthy TV consumption) is critically evaluated. When designing PT, three design areas should be considered "autonomy and free choice", "coercion versus reflection", and "surveillance and privacy". Verbeek [55] researched the perils of ambient intelligence and PT. Technology is not only a neutral enabler of behaviors, but it also shapes how people act and experience reality. Thus, PT requires reconsidering the concept of human freedom and our understanding of both moral and casual responsibility. The author elaborates on the responsibility of both the user and the designer.

Another area of concerns is about what in PT is morally and ethically acceptable [24, 25, 29, 43, 49]. Guerini et al. [24] have researched the moral implications and actions of autonomous artificial agents, e.g., adaptive PT. The authors emphasize that flexibility is important for PT through adapting a persuasive strategy to fit the situation and the character of the persuaded. Ham and Spahn [25] looked at the physiological effects and moral acceptability of persuasive robots. An important issue to consider is alternative persuasive strategies and what means to reach the aim. The importance to identify persuasive principles and attention to ethical consideration is emphasized. Page and Kray [43] used focus groups to understand relevant ethical aspects of PT in the context of healthy living. The result showed three factors that people value when determining the ethical acceptability of PT, namely the commissioner, the recipient, and the means of delivery. Text messages were seen as more acceptable and electrical shock or bank account restrictions most unethical. Interesting to notice is that electric shock could sometimes be justified. Stock et al. [49] researched adaptive PT to gain better understanding of the moral acceptability of the communicative action conducted by PT to reach its goal of persuading. One interesting finding is that people do not seem to evaluate the moral acceptability of machine persuasion differently compared to human persuasion, despite the fact that a priori most answered that machine persuasion could not be morally accepted. The persuasive system should be flexible with persuasion strategies and adaptive to the persuaded. Kaptein and Eckles [29] showed concerns regarding adaptive persuasive systems, as they rarely disclose the system's ability to adapt to individual differences and that a system trained in one context could be used in other unexpected ways. The systems could create persuasion profiling of an individual, and this may become ethically challenging, as the personal data could be distributed and shared between systems without any consent from the user.

Privacy of PT is also something that needs attention [6, 35]. Leth et al. [35] showed valid ethical concerns that PT could contribute to the surveillance of individuals. The authors discuss how Berdichevsky and Neuenschwander's framework and Fogg's stakeholder analysis could be used as a help to ethical problems, because many systems could be used for surveillance, depending on the context, and for some this possibility might be quite tempting. The persuasive system should not violate individual privacy.

Researchers [45, 53] are also interested in the volunteerism of PT . Timmer et al. [53] wrote about an important ethical issues for PT, e.g., persuasive systems that are used at work could, depending on context and viewpoint, be seen as

mandatory. Thus, the system use would not be perceived as voluntary, and group pressure at work could influence users. Ethically responsive PT should preserve the autonomy of an individual, and this is something the designer of PT must consider.

Another theme that concerns the design of PT, is about guidelines and evaluation [6, 36, 45, 47, 56]. Spahn [47] and Linder [36] both apply discourse ethics to PT as a way to understand the ethics of this technology. Spahn derives various criteria from discourse ethics for usage and design of PT, so these criteria could be used as a guideline. Linder elaborates on the assessment of PT, as it is a medium for the designer, the engineer, or authority to change the user's behavior towards planned goals. The principals of discourse ethics could be a way to reflect PT, but Linder also demonstrates the limitation of discourse ethics. Yetim [56] use value-based practical reasoning and argumentation schemes as a foundation to build a framework for practical discourse. The questions are remapped into practical-, ethical-, and moral discourse and could be used when designing, evaluating, and critically assesses the goals, values, and actions of a persuasive system. Stibe and Cugelman [48] have demonstrated how PT could backfire and calls for a discussion concerning negative outcomes of PT. To aid this discussion, they have developed the "Intention-Outcome" and risk managing "Likelihood-Severity" matrices, as well as a taxonomy for categorizing persuasive backfire. de Oliveira and Carrascal [42] proposed new approaches when designing PT to highlight necessary ethical concerns and to wake the awareness of both designers and users of PT. They explored three approaches: an enforced prevention (e.g. guidelines provided by government or organizations), an encouraged prevention (e.g. voluntary certification), and a remediation-based approach (e.g. tools for users to reveal, identify, and remove or mitigate the bias of PT).

2.3 Dark patterns and computer game

Design pattern as a concept was introduced by Christopher Alexander as a solution that is proven and reusable for an architectural design problem [2]. Design patterns have, for example, been used in interaction design [10], software engineering [21], and game design [8], as a reusable solution for problems in a specific context. A pattern solution often captures more solutions in preference of one exact solution. Dark patterns were introduced by Harry Brignull when he cataloged (on darkpatterns.org) different types of interfaces that trick users into doing things that are not in their best interest [11].

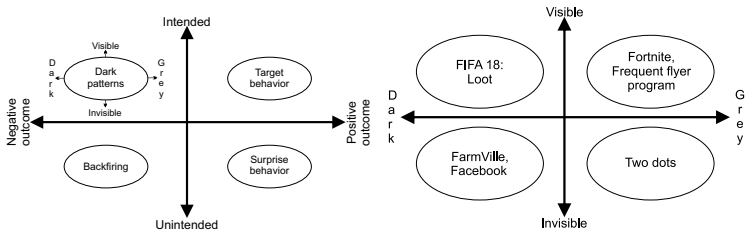
Hence, a dark pattern design could be defined as: *the craft of purposefully designing patterns that damage the well-being of the users.*

Related to that, Zagal et al. did research on dark game design patterns, which can be seen as unethical and questionable [58]. Linehan et al. developed dark design patterns for anti-health games [37]. The negative experiences for players are likely to happen without their consent and against their best interest. Dark patterns are design strategies that are used to benefit developers more than the target audience, e.g., using unethical applications, such as coercion, deception,

and fraud. Any design pattern becomes dark at the moment when it intentionally unbalances the well-being gains towards the creator of Pt and away from its users.

3 Visibility-Darkness matrix

To explore the dark side of behavioral designs, Stibe and Cugelman [48] have introduced the Intention-Outcome matrix that has four quadrants. Target behavior is the primary intended positive behavioral outcome being designed for. Surprise behavior is a positive behavioral outcome that was not intended, however is a complementary benefit of the behavioral design that contributes to the well-being of users. Backfiring includes several negative outcomes, like a side effects, when the target behavior is achieved, but the design also triggers unintended negative outcomes.



(a) The Intention-Outcome matrix (extended from [48]).

(b) Visibility-Darkness matrix with examples.

Fig. 1: Matrices of darkness

As an extension of the Intention-Outcome matrix (Figure 1a), we propose to uncover the deeper dimensions for classifying dark patterns, which is the fourth quadrant (upper left) of the Intention-Outcome matrix. Dark patterns often are made invisible to the users of influencing systems. For example, websites can hide their true intentions of why they collect user information. Sometimes, partial information is made available in small print way down the structure of a website. Such approach makes dark patters to be less visible to the users. In a few occasions, some of the information describing a dark pattern is made quite visible to the users, however many of them tend to be careless while rushing through their chosen PT, so they pay little to none attention. Dark patters also can be of various intensity, meaning they can have different sizes of impact on users and their desired outcomes while interacting with PT. Some of the dark patterns may leave a very small footprint on user experience, while others can seriously challenge personal well-being of individual users. There are persuasive designs that tend to increase addiction, for example, which is a very dark pattern as such. Other interaction design patterns may not be that dark, for example, only collecting user information and then emailing updates to subscribers without their proper consent. That still is a dark pattern in the shade of grey, but not pitch-dark.

Thus, we propose subdividing the dark patterns quadrant into a matrix that contrasts the visibility of dark patterns (visible to invisible) with the shade of darkness (grey to dark). Based on that, we introduce the Visibility-Darkness matrix Figure 1b. Further, we provide and discuss examples as illustrations for the four quadrants of the Visibility-Darkness matrix.

Visible-Grey quadrant: Here we have PT designed to be beneficial for the users, but an outcome may not be as good as it is presented for the individuals. In other words, the potential benefits are clearly emphasized, while all potentially inconvenient extras are given as undebatable. Thus, everything looks like to be visible, however there seems to be an unfair divide of gains between the designer and the users. For example, different bonus systems, such as frequent flyer miles and alike. Users may often get manipulated into buying more products or service than necessary. Traditional bonus systems usually rely on badges and leaderboards, i.e. different levels that give benefits and increase status (upgrades, lounges). Participants at times need to spend a certain amount of money to keep their current level. Also, the previously earned bonus points may have an expire date that would clearly encourage users to buy, rent, or fly extra to keep the points Fortnite is a free-to-play video game by Epic Games that became a viral phenomenon in year 2018. The in-game store offers outfit customization, dance moves, etc. to make the game more fun to play. Fortnite's in-game currency is V-bucks (1000 V-buck cost US $9.99). V-bucks can also be earned through completing game missions. The customization does not bring any competitive advantage against other players. The game developers are regularly introducing new game enhancements, persuading people to continue playing. Many players are young and feel persuaded to buy the same things as their friends, therefore the total amount of money spent on the game after a while can become high. Although, the game brings enjoyment to players, the outcome may not be always that beneficial for them. For example, a sort of game addiction may emerge, as well as a form of coercion for parents to buy V-bucks.

Invisible-Grey quadrant: Here we have PT designed with features that may not be clearly seen or properly understood by the users. At the same time, such implementations by definition are also aimed at bringing more benefits to the designers versus the users. In other words, not only the gains are skewed towards the designers, but they also try hiding their intentions under the surface of misleading user interaction. An example here is a mobile game called Two Dots that is developed by PlayDots and available on Google Play and App Store. The game is quite minimalistic in design. When a user loses a game, pressing a green button usually means to continue. However, once all the available lives are lost, the user is seeing a familiar green box, but now it means to pay US $ 0.99 to continue [54]. This could be classified as a learned conditioned stimulus, when users press the green button by a reflex to continue the game despite a small "x" is available for canceling that action.

Invisible-Dark quadrant: Here we have PT designed with an intention that is not clearly visible, as well as the potential damage to the users may be quite large. This quadrant is actually the place, where the darkest patterns can hide. Because they can be very dark and well camouflaged at the same time. By definition, here the designers

would be abusing the weaknesses of human nature. Zynga is a game company that produced Facebook games like FarmVille, allowing players to use microtransactions for buying in-game benefits. These kind of Facebook games has been criticized for being designed for revenue and multiply users in every possible way [20]. Ian Bogost created a game "Cow Clicker" that mimics the social games on Facebook in a satiric way. He criticizes the compulsive and time destroying elements of such games [9].

Facebook uses confidential algorithms for persuading users to read and interact with their news feed on the platform [14]. The algorithm that recommends news could act in Facebook's best interest and not the user's well-being. Bessi et al. [7] found that Facebook users typically engage with information that confirms with their thinking. That could increase the chance of addiction to the social media site. This gives rise for a new relationship between machines and humans transforming the prioritization of news and their interpretation [14].

Visible-Dark quadrant: Here we have PT designed with an intention that can look dubious to the users. Electronic Arts is the maker of FIFA 18 a football game, where players can buy "Loot packs" to increase their chance to beat opponents. The loot pack's content is randomized, many game players say they need to buy loot packs to stay competitive. They claim that the game design is "pay-to-win" and unfair. Some people are also suffering from game addiction similar to gambling, by spending much money on loot packs. Ther is an ongoing debate in Sweden about the ethics and the mechanics of the game, as it is almost coercive to buy the loot packs [51, 52]. FIFA 18 is not unique, there are other games where loots are used: Valve's "Counter-Strike: Global Offensive (CS GO)", and Blizzard's "Overwatch". CS Go have Lootmarket.com where players can buy items from loots.

Popularity have gained free games and apps with inbuilt "motivation" to pay for premium content later. A Swedish newspaper reported that children were able (without security codes) to buy in-game items for US $ 5550 during one month by playing The Smurfs' village [1]. There are numerous free games that allow players to buy in-games item to boost performance. If friends are buy boosts, social dynamics can motivate others to buy boosts. An example of such a game is Candy Crush Saga, with items to unlock next level, or to boost gameplay. After losing lives, instead of waiting a certain time to get a new life the player can pay to continue playing. The game players may have invested a lot of time in the game and hence have inner incentives and motivation to pay the fee to continue. Some game levels are extremely difficult, so the game could be designed with an intention that the players have to buy in-game items in order to enjoy the game and keep up with other players.

Evaluation of the darkness of PT Fogg [19] recommended 7 steps that designers can use to evaluate the ethical nature of PT by its outcome, methods, and intentions. Berdichevsky and Neuenschwander [6] developed a framework to evaluate the ethics of PT. It's focus is on interaction between persuader, PT, and persuaded person. Because the persuader designs and creates PT, which can be seen as a technical mediation, see Latour [32], that uses persuasive methods on the persuaded person, resulting in an outcome (both predictable and unpredictable). Thus, we adapted this framework for evaluation of the PT darkness (see Figure 2).

The designer of PT creates an experience with a set goal. PT uses different game mechanics depending on the context, e.g., resources, feasibility and time frame, etc. The well-being is the outcome of PT that constitute a benchmark for the evaluation using the Visibility-Darkness matrix. It is important to not regard technology as a neutral instrument, because, in a social context, it is value-laden [50], and aims at transforming user's behavior.

Fig. 2: Framework for Evaluating the Darkness of Persuasive Technology (FEDPT).

Below the FEDPT is used to elaborate on the previous given examples for the Visibility-Darkness matrix:

Visible-Grey - Well-being: The users think something is beneficial for them, however may end up with something that is not in their best interest. PT may have clear visible rules, but a holistic result and the well-being impact for the user is difficult to foreseen completely. The users may think they have chosen the best option but most benefits might go to the PT owner.

Invisible-Grey - Well-being: In this case, the PT design is aimed at getting users hooked and react to stimulus in a certain way, without providing all relevant information appropriate. Later, this is used in making people follow a learnt behavior, which may not bring any well-being to the user. The PT designer has implemented options and the user can try complaining, but might not do so if the micro-transaction has a low perceived value. The PT designer takes a calculated risk on users willingness to recover their loss.

Invisible-Dark - Well-being: The user may not really understand possible outcomes or purpose of PT, as it seems to be a repetitive and never-ending game. The user could get hooked by friends into using social games with no real challenge (just wait and click), and then waste time and perhaps money by purchasing all in-game items. With an aim of increasing interaction and revenue, hidden algorithms can persuade and change the way users interact on a social network site. The algorithm could transform how the users are interpreting news. A PT designer may optimize their news and increase their impact, the logic is hidden from the users so they are left to the grace of PT designers.

Visible-Dark - Well-being: Although, the user can see and understand the purpose of PT, they still can be easily hooked and get addicted. After a conscious reflection that actually may feel as a manipulation into "wasting" more time and money. This should be especially well controlled and monitored for PT aimed at children, where the designers should have an even larger responsibility, therefore necessary to aim for the highest degree of trustworthiness.

4 Use case

The FEDPT is validated thru use case methodology. Use case is a method that gives a foundation for higher-level verification. The interaction sequences between users of a system and the system related to a specific goal is represented through the use cases. One research article that reflects similar system design is selected for the validation as a use case can show possible system activities in the interaction between a system and users.

The persuasive system design (PSD) model have four persuasive system principles: primary task support, dialogue system support, perceived credibility, and social support [41]. Case studies should focus on contemporary issues in real life and be grounded on the managerial or organizational level [57]. Using a use case conforms to the purpose of using case studies in qualitative research where it has been argued that case studies could be used to test theory within the positivist paradigm [15, 33, 34] and to synthesize insight from previous research what Seddon and Scheepers calls theory building [46]. Support carrying out of the user's primary task principles includes reduction, tunneling, personalization, tailoring, self-monitoring, simulation, and rehearsal. System principles to support implementing computer-human dialogue includes liking, praise, rewards, reminders, suggestion, similarity, and social role. System principles that gives system credibility consists of credibility, trustworthiness, surface, real-world feel, expertise, authority, third-party endorsements, and verifiability. And finally, the system principles belonging in the social support category are social comparison, social facilitation, normative influence, competition, social learning, cooperation, and recognition. The use-case diagram of FEDPT mapped with PSD is shown in Figure 3. PSD is crucial and chosen for the use case since it could be considered fundamental in designing and evaluating persuasive systems and PT. The PSD in the context of FEDPT gives a clearer awareness for the designer of the goal of wellness and the possible dangers of dark and unintended outcomes. The actor in the use case aims towards well-being goals for the persuasive system.

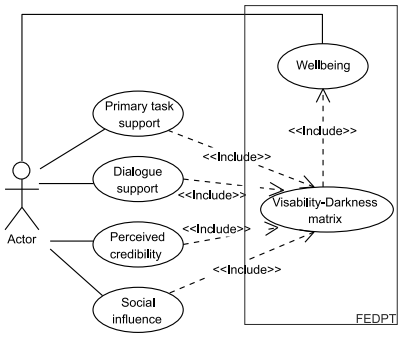


Fig. 3: Use case of PSD with FEDPT.

5 Discussion

When taking a closer look into the dark side of PT, many patterns become recognizable, as they make people addicted to what they use or game experiences they have. Although, the creators of such PT would argue that their designs are intended to keep users engaged, in many cases it goes beyond what might be perceived as positive contributions to well-being. There must be a clearer way for everyone to see and recognize how their PT engagements interplay with their own well-being. One could suggest John Rawl's "veil of ignorance" [44], hence the designers should not design PT that they would not use themselves. That is

similar to Berdichevsky and Neuenschwander's "the golden rule of persuasion", the designer should not create PT that persuade to something they themselves would not like to be persuaded to do. Ultimately, such requirements and vision shall be a prerequisite for the PT scholars and professionals, so incorporated as an essential part of their design and development processes. Research has shown how UX design practitioners are tempted by the dark (pattern) side to implement dubious design [23].

More than ever before, it now becomes very important to engage PT designers to create a common understanding of how the essentials of their work are influencing and determining the lives of millions. The scientific work that we have outlined in this paper contributes to sharper understanding of how the design and strategical choices of the PT experience can potentially appear to be harmful for the well-being of users. The FEDPT, including the Visibility-Darkness matrix, shall now be very instrumental for many PT scholars and practitioners to assess their designs and evaluate possible negative effects on the users and the overall user experience. Although, it could be argued that a sinister designer of dark PT could use the FEDPT to enhance the darkness, the FEDPT reveal and makes the dark patterns a shared knowledge. All the stakeholders of PT, i.e. scholars, users, designers, professionals, etc. can now use FEDPT to evaluate PT and perhaps also certify the persuasive user experiences as not having a negative impact on well-being.

Particular care should be taken when designing PT for children. There should be requirements of extra safety mechanism for in-game purchases (something that Alha et al. [3] noticed for free-to-play games). The focus on extrinsic rewards should be kept at a minimum, as such rewards are often useless in the real life, regardless fo their ability to boost self-confidence or perceived status. Also devices get increasingly connected, the Internet of things (IoT), thus the designers should be careful when designing PT that takes advantage of these new capabilities such as continuously collecting data in a secure manner and respecting the right to privacy. PT such as gamification is popular so companies unfortunately started implementing it without properly understanding the essence of game mechanics, flow, immersion, and story. That often can result in PT that does not go in line with the intended positive goal or result in negative outcomes.

6 Conclusions and future research

A proper and meaningful discussion around and research on the negative consequences of PT and behavioral change designs has now become inevitable, when thinking and caring about our collective future well-being. Many scholars and practitioners have already raised related concerns over the last years. More importantly, such debate must be an integral part of all efforts aimed at designing technology for influencing human behavior. Scholars are now encouraged to co-create new scientific knowledge by using, applying, and extending the proposed FEDPT framework and the Visibility-Darkness matrix. Practitioners are urged

to include this work into their daily processes for designing PT and delivering products and experiences. Particularly, we invite interested researchers and practitioners of PT to join our work by providing additional examples of dark patterns. There is a need to continue exploring and monitor emerging forms of dark PT, so to advance the knowledge and refining the proposed framework. For example, we need a clearer discussion on ways for differentiating examples of visible versus invisible and grey versus dark patterns. Having such fundamental work progressing, we shall be able to create a taxonomy of dark patterns, including sharper guidelines for classifying designs that produce negative outcomes.

References

1. Aftonbladet: Barnens ipad spel kostade 50 000 kr (2011), retrieved February 7, 2018 from https://www.aftonbladet.se/nyheter/article12846738.ab
2. Alexander, C.: A pattern language: towns, buildings, construction. Oxford university press (1977)
3. Alha, K., Koskinen, E., Paavilainen, J., Hamari, J., Kinnunen, J.: Free-to-play games: Professionals' perspectives. Proceedings of nordic digra 2014 (2014)
4. Andrade, F.R.H., Mizoguchi, R., Isotani, S.: The bright and dark sides of gamification. In: Micarelli, A., Stamper, J., Panourgia, K. (eds.) Intelligent Tutoring Systems. pp. 176–186. Springer International Publishing, Cham (2016). https://doi.org/10.1007/978-3-319-39583-8_17
5. Atkinson, B.M.C.: Captology: A critical review. In: IJsselsteijn, W.A., de Kort, Y.A.W., Midden, C., Eggen, B., van den Hoven, E. (eds.) Persuasive Technology. pp. 171–182. Springer Berlin Heidelberg, Berlin, Heidelberg (2006). https://doi.org/10.1007/11755494_25
6. Berdichevsky, D., Neuenschwander, E.: Toward an ethics of persuasive technology. Communications of the ACM **42**(5), 51–58 (1999). https://doi.org/10.1145/301353.301410
7. Bessi, A., Zollo, F., Del Vicario, M., Puliga, M., Scala, A., Caldarelli, G., Uzzi, B., Quattrociocchi, W.: Users polarization on facebook and youtube. PloS one **11**(8), e0159641 (2016). https://doi.org/10.1371/journal.pone.0159641
8. Björk, S., Holopainen, J.: Patterns in Game Design (Game Development Series). Charles River Media, Inc., Rockland, MA, USA (2004)
9. Bogost, I.: Blog: Cow clicker - the making of obsession (2018), retrieved February 10, 2018 from http://bogost.com/blog/cow_clicker_1/
10. Borchers, J.O.: A pattern approach to interaction design. AI & SOCIETY **15**(4), 359–376 (Dec 2001). https://doi.org/10.1007/BF01206115
11. Brignull, H.: Darkpatterns.org (2018), retrieved April 9, 2018 from https://darkpatterns.org
12. Callan, R.C., Bauer, K.N., Landers, R.N.: How to avoid the dark side of gamification: Ten business scenarios and their unintended consequences. In: Reiners, T., Wood, L.C. (eds.) Gamification in Education and Business. pp. 553–568. Springer International Publishing, Cham (2015). https://doi.org/10.1007/978-3-319-10208-5_28
13. Deterding, S., Dixon, D., Khaled, R., Nacke, L.: From game design elements to gamefulness: Defining "gamification". In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. pp. 9–15. MindTrek '11, ACM, New York, NY, USA (2011). https://doi.org/10.1145/2181037.2181040

14. DeVito, M.A.: From editors to algorithms. Digital Journalism **5**(6), 753–773 (2017). https://doi.org/10.1080/21670811.2016.1178592
15. Eisenhardt, K.M.: Building theories from case study research. Academy of management review **14**(4), 532–550 (1989). https://doi.org/10.5465/amr.1989.4308385
16. Etkin, J.: The hidden cost of personal quantification. Journal of Consumer Research **42**(6), 967–984 (2016). https://doi.org/10.1093/jcr/ucv095
17. Fishbach, A., Choi, J.: When thinking about goals undermines goal pursuit. Organizational Behavior and Human Decision Processes **118**(2), 99–107 (2012). https://doi.org/10.1016/j.obhdp.2012.02.003
18. Fogg, B.: Persuasive computers: Perspectives and research directions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 225–232. CHI '98, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA (1998). https://doi.org/10.1145/274644.274677
19. Fogg, B.: Persuasive Technology: Using Computers to Change what We Think and Do. Morgan Kaufmann (2003)
20. Gamasutra: Zynga: The future, or just a bit of it? (2010), retrieved February 12, 2018 from https://www.gamasutra.com/blogs/DavidHayward/20100315/4670/Zynga_The_Future_Or_Just_A_Bit_Of_It.php
21. Gamma, E., Helm, R., Johnson, R., Vlissides, J.: Design patterns: elements of reusable object-oriented software. Addison-Wesley (1995)
22. Gram-Hansen, S.B.: Persuasive everyware-possibilities and limitations. In: 14th World Multi-Conference on Systemics, Cybernetics and Informatics: WMSCI 2010. pp. 254–260. International Institute of Informatics and Systemics (2010)
23. Gray, C.M., Kou, Y., Battles, B., Hoggatt, J., Toombs, A.L.: The dark (patterns) side of ux design. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. pp. 534:1–534:14. CHI '18, ACM, New York, NY, USA (2018). https://doi.org/10.1145/3173574.3174108
24. Guerini, M., Pianesi, F., Stock, O.: Is it morally acceptable for a system to lie to persuade me? In: Artificial Intelligence and Ethics: Papers from the 2015 AAAI Workshop. vol. WS-15-02, pp. 53–60 (2015)
25. Ham, J., Spahn, A.: Shall i show you some other shirts too? the psychology and ethics of persuasive robots. In: Trappl, R. (ed.) A Construction Manual for Robots' Ethical Systems: Requirements, Methods, Implementations, pp. 63–81. Springer International Publishing, Cham (2015). https://doi.org/10.1007/978-3-319-21548-8_4
26. Hamari, J., Koivisto, J., Pakkanen, T.: Do persuasive technologies persuade? - a review of empirical studies. In: Spagnolli, A., Chittaro, L., Gamberini, L. (eds.) Persuasive Technology. pp. 118–136. Springer International Publishing, Cham (2014). https://doi.org/10.1007/978-3-319-07127-5_11
27. Huotari, K., Hamari, J.: Defining gamification: A service marketing perspective. In: Proceeding of the 16th International Academic MindTrek Conference. pp. 17–22. MindTrek '12, ACM, New York, NY, USA (2012). https://doi.org/10.1145/2393132.2393137
28. Hyrynsalmi, S., Smed, J., Kimppa, K.K.: The dark side of gamification: How we should stop worrying and study also the negative impacts of bringing game design elements to everywhere. In: Proceedings of the 1st International GamiFIN Conference. pp. 96–104. CEUR Workshop Proceedings (2017)
29. Kaptein, M., Eckles, D.: Selecting effective means to any end: Futures and ethics of persuasion profiling. In: Ploug, T., Hasle, P., Oinas-Kukkonen, H. (eds.) Persuasive Technology. pp. 82–93. Springer Berlin Heidelberg, Berlin, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13226-1_10

30. Kim, T.W., Werbach, K.: More than just a game: Ethical issues in gamification. Ethics and Information Technology **18**(2), 157–173 (Jun 2016). https://doi.org/10.1007/s10676-016-9401-5
31. Kuonanoja, L., Oinas-Kukkonen, H.: Recognizing and mitigating the negative effects of information technology use: A systematic review of persuasive characteristics in information systems. In: Müller, S.D., Nielsen, J.A. (eds.) Nordic Contributions in IS Research. pp. 14–25. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-319-96367-9_2
32. Latour, B.: On technical mediation. Common knowledge **3**(2), 29–64 (1994)
33. Lee, A.S.: A scientific methodology for mis case studies. MIS quarterly pp. 33–50 (1989)
34. Lee, A.S., Baskerville, R.L.: Generalizing generalizability in information systems research. Information systems research **14**(3), 221–243 (2003). https://doi.org/10.1287/isre.14.3.221.16560
35. Leth Jespersen, J., Albrechtslund, A., Øhrstrøm, P., Hasle, P., Albretsen, J.: Surveillance, persuasion, and panopticon. In: de Kort, Y., IJsselsteijn, W., Midden, C., Eggen, B., Fogg, B.J. (eds.) Persuasive Technology. pp. 109–120. Springer Berlin Heidelberg, Berlin, Heidelberg (2007). https://doi.org/10.1007/978-3-540-77006-0_15
36. Linder, C.: Are persuasive technologies really able to communicate?: Some remarks to the application of discourse ethics. International Journal of Technoethics (IJT) **5**(1), 44–58 (2014). https://doi.org/10.4018/ijt.2014010104
37. Linehan, C., Harrer, S., Kirman, B., Lawson, S., Carter, M.: Games against health: A player-centered design philosophy. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems. pp. 589–600. CHI EA '15, ACM, New York, NY, USA (2015). https://doi.org/10.1145/2702613.2732514
38. Lupton, D.: Self-tracking modes: Reflexive self-monitoring and data practices. In: Imminent Citizenships: Personhood and Identity Politics in the Informatic Age - Workshop. SSRN (2014). https://doi.org/10.2139/ssrn.2483549
39. Nyström, T.: Gamification of persuasive systems for sustainability. In: 2017 Sustainable Internet and ICT for Sustainability (SustainIT). IEEE (2017). https://doi.org/10.23919/SustainIT.2017.8379815
40. Oinas-Kukkonen, H.: A foundation for the study of behavior change support systems. Personal and Ubiquitous Computing **17**(6), 1223–1235 (Aug 2013). https://doi.org/10.1007/s00779-012-0591-5
41. Oinas-Kukkonen, H., Harjumaa, M.: Persuasive systems design: Key issues, process model, and system features. Communications of the Association for Information Systems **24**, 28 (2009). https://doi.org/10.17705/1CAIS.02428
42. de Oliveira, R., Carrascal, J.P.: Towards effective ethical behavior design. In: CHI '14 Extended Abstracts on Human Factors in Computing Systems. pp. 2149–2154. CHI EA '14, ACM, New York, NY, USA (2014). https://doi.org/10.1145/2559206.2581182
43. Page, R.E., Kray, C.: Ethics and persuasive technology: An exploratory study in the context of healthy living. In: Proceedings of the First International Workshop on Nudge & Influence Through Mobile Devices. vol. 690, pp. 19–22. CEUR-WS (2010)
44. Rawls, J.: A theory of justice: Revised edition. Harvard university press (1999)
45. Reitberger, W., Güldenpfennig, F., Fitzpatrick, G.: Persuasive technology considered harmful? an exploration of design concerns through the tv companion.

In: Bang, M., Ragnemalm, E.L. (eds.) Persuasive Technology. Design for Health and Safety. pp. 239–250. Springer Berlin Heidelberg, Berlin, Heidelberg (2012). https://doi.org/10.1007/978-3-642-31037-9_21
46. Seddon, P.B., Scheepers, R.: Generalization in is research: a critique of the conflicting positions of lee & baskerville and tsang & williams. Journal of Information Technology **30**(1), 30–43 (2015). https://doi.org/10.1057/jit.2014.33
47. Spahn, A.: And lead us (not) into persuasion...? persuasive technology and the ethics of communication. Science and Engineering Ethics **18**(4), 633–650 (Dec 2012). https://doi.org/10.1007/s11948-011-9278-y
48. Stibe, A., Cugelman, B.: Persuasive backfiring: When behavior change interventions trigger unintended negative outcomes. In: Meschtscherjakov, A., De Ruyter, B., Fuchsberger, V., Murer, M., Tscheligi, M. (eds.) Persuasive Technology. pp. 65–77. Springer International Publishing, Cham (2016). https://doi.org/10.1007/978-3-319-31510-2_6
49. Stock, O., Guerini, M., Pianesi, F.: Ethical dilemmas for adaptive persuasion systems. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. pp. 4157–5161. AAAI'16, AAAI Press (2016)
50. Sundström, P.: Interpreting the notion that technology is value-neutral. Medicine, Health Care and Philosophy **1**(1), 41–45 (Apr 1998). https://doi.org/10.1023/A:1009933805126
51. Sveriges Radio: Sr 3 - radio news: Fifa 18 loot packs part 1 (2018), retrieved February 7, 2018 from http://sverigesradio.se/sida/avsnitt/1031434?programid=1646
52. Sveriges Radio: Sr 3 - radio news: Fifa 18 loot packs part 2 (2018), retrieved February 8, 2018 from http://sverigesradio.se/sida/avsnitt/1034069?programid=1646
53. Timmer, J., Kool, L., van Est, R.: Ethical challenges in emerging applications of persuasive technology. In: MacTavish, T., Basapur, S. (eds.) Persuasive Technology. pp. 196–201. Springer International Publishing, Cham (2015). https://doi.org/10.1007/978-3-319-20306-5_18
54. User Testing Blog: Dark patterns: The sinister side of ux (2015), retrieved February 10, 2018 from https://www.usertesting.com/blog/2015/10/01/dark-patterns-the-sinister-side-of-ux/
55. Verbeek, P.P.: Ambient intelligence and persuasive technology: The blurring boundaries between human and technology. NanoEthics **3**(3), 231 (Dec 2009). https://doi.org/10.1007/s11569-009-0077-8
56. Yetim, F.: A set of critical heuristics for value sensitive designers and users of persuasive systems. In: ECIS 2011 (2011)
57. Yin, R.K.: Case study research: Design and methods. Sage publications, sixth edn. (2017)
58. Zagal, J.P., Björk, S., Lewis, C.: Dark patterns in the design of games. In: Proceedings of the Conference on Foundations of Digital Games 2013 (2013)