Gray Areas between Persuasion and Manipulation in UX Design: An Exploratory Study in Japan

Naomi V Panjaitan
Faculty of Science and Engineering
Waseda University
Tokyo, Japan
naomivp@toki.waseda.jp

Katsumi Watanabe
Faculty of Science and Engineering
Waseda University
Tokyo, Japan
katz@waseda.jp

***Abstract*—** The alignment—or misalignment—of design intent and its impact on the user is a critical concern in the field of user experience (UX) design, particularly as the distinction between persuasion and manipulation becomes increasingly blurred. Our study, conducted with 353 participants in Japan, investigated users' ability to detect the manipulative nature of deceptive design patterns. Our findings indicate that while some dark patterns are easily recognized as manipulative by users, others are more subtle and may not be perceived as manipulative at all. Conversely, an unexpected result of our research was that manipulative intent was perceived in three out of four persuasive designs that were not inherently deceptive. Through this exploration, we have highlighted the critical role of human cognition and inherent biases in shaping user experiences, and the imperative for UX professionals to navigate these complexities with caution. We also discussed recommendations for UX practitioners on proactive steps to address the uncertainty in distinguishing between manipulative and persuasive designs.

Keywords—dark pattern, deceptive design, UX design, cognitive bias

I. Introduction

In the digital age, every click and swipe holds significance. Businesses adeptly employ digital nudges to subtly influence user behavior [1] and encourage actions like purchases, subscriptions, or extended engagement on their platforms.

Fundamentally, digital nudges are designed to be persuasive, engaging directly with an individual's ability to make conscious decisions and choices through clear reasoning or presented incentives [1], [2]. However, these nudges run the risk of becoming manipulative when they are concealed or take advantage of vulnerabilities in the decision-making process [2]. For instance, the use of 'dark patterns' in digital design.

Dark patterns, also known as 'deceptive patterns,' are design techniques that take advantage of users' cognitive biases to manipulate them into taking actions that benefit the company, often at the user's expense [3], [4]. For example, the "hidden cost" strategy adds unexpected charges at the end of a transaction [4]. This tactic exploits the sunk cost fallacy [5], where people continue a purchase to justify the time or effort they've already invested.

In recent years, dark patterns have gained increasing attention, with researchers developing taxonomies [3], [4], [6] for systematic classification and identification of dark patterns and raising interest from various stakeholders worldwide, such as the FTC [7] and EDBP [8]. In response to this growing international focus, our research endeavors to explore the landscape of dark patterns within Japan.

In our previous study [9], we engaged Japanese participants in a comprehensive study designed to understand how users perceive and identify these deceptive designs. We reported on the general patterns and tendencies in how Japanese people respond to dark patterns.

In the present study, we conducted additional signal detection analyses and found that certain dark patterns are overwhelmingly easier to detect, while others are more difficult to detect. Interestingly, our analyses also revealed instances where persuasive designs that are not traditionally considered deceptive are deemed manipulative by participants, highlighting the nuanced understanding and awareness of dark patterns among users.

II. Literature Review

A. Deceptive Pattern Taxonomies

Researchers have developed various taxonomies to dissect and categorize these deceptive design strategies. Among the most notable contributions to the field are the taxonomies developed by Harry Brignull, Colin M. Gray et al., and Arunesh Mathur et al.

Harry Brignull, a pioneer in the study of dark patterns, coined the term "dark pattern" in 2010. His taxonomy was derived from personal observations and reports of tricks used in websites and apps to get users to do things they did not mean to do, such as buy or sign up for something unintentionally [4]. Brignull's work is considered a cornerstone in the field of deceptive design.

Following this, Gray et al. proposed to organize existing dark patterns into five major categories in their 2018 paper. These categories include: Nagging, Obstruction, Sneaking, Interface Interference, and Forced Action [6]. This paper provided a solid foundation for identifying recurring dark patterns across domains.

Subsequently, Arunesh Mathur and colleagues conducted a comprehensive study in 2019, which involved crawling over 11,000 shopping websites to identify and categorize dark patterns [3]. Their research identified 15 types and 7 broader categories of dark patterns, based on their characteristics and the cognitive biases they exploit in users.

B. Decisions and Behavior

Amos Tversky and Daniel Kahneman's work on heuristics and biases shows that under cognitive stress, such as information overload or time pressure, we tend to fall back on these mental shortcuts and can easily deviate from rational judgment [10]. The effectiveness of deceptive designs is

This work was supported by KAKENHI 22H00090. The authors have no conflicts of interest to declare that are relevant to the content of this article.

rooted in psychological principles—they exploit common heuristics and biases that affect decision-making:

- Default bias, which exploits our inclination to stick with predefined options, capitalizes on users' general tendency to avoid the effort required to opt out [11]. Consequently, companies might implement pre-checked boxes [3], [4] that steer users toward opt-into choices that are beneficial (e.g., subscribing to a newsletter, agreeing to a data collection, etc.).
- Scarcity bias is leveraged by displaying false alerts about dwindling stock, forcing customers to make hasty decisions out of fear of missing out [2]. This approach preys on our tendency to value items that are perceived as rare or in short supply, often leading to rash purchase decisions [12].
- The sunk cost fallacy exploits our propensity to stick with an endeavor after investing significant resources [5]. The 'hidden cost' dark pattern occurs when companies conceal additional fees or charges until deep into the transaction process [2]. This pattern banks on our reluctance to forgo investments of time, effort, or money, often resulting in acquiescence to unexpected costs or unfavorable terms.

C. Persuasive Technology and Nudge

Building on the understanding from the prior section about cognitive loads, BJ Fogg, in his book "Persuasive Technology: Using Computers to Change What We Think and Do" suggests that in order to induce behavioral change, it is effective to offer 'hot triggers' to motivated individuals. He suggests a three-step process for designing the technology to influence behavior: be specific, make it easy, and trigger behavior [13].

Nudges could serve as 'hot triggers' in Fogg's model, acting as specific catalysts in the presence of motivation to prompt the desired behavior. The concept of nudges—as introduced by Richard H. Thaler and Cass R. Sunstein, is a subtle design features that guide users toward certain decisions without restricting choice [1].

At their core, both persuasive technologies and nudges leave the final decision to the user. However, these strategies veer toward manipulation when they compromise user autonomy [14] and obscure the clarity [2], [15] needed to make informed decisions.

III. Survey

A. Methodology

We recruited 353 participants through Yahoo Crowdsourcing, a platform that connects businesses with freelancers for various tasks. Participants were directed to the Qualtrics platform for survey completion and data gathering. We utilized MathWorks MATLAB to organize, clean, and perform statistical analysis on the data.

The authors used the aforementioned dataset collected in another study to gauge individual differences (i.e., demographics and personality traits) in the ability to detect the manipulative nature of deceptive patterns [9]. The present study is entirely different in scope and focus, providing new insights and analyses into the manipulative nature of dark patterns.

Beyond individual differences in perception, this study aimed to explore the nuanced differences between persuasive and manipulative design elements. Specifically, it examined why certain dark patterns are more easily identified than others, and why some non-dark patterns are perceived as manipulative by users in a case-study fashion.

In the survey, we showcased four dark patterns identified in various established taxonomies [3], [4], [6] to ensure a broad representation of deceptive designs. The selected dark patterns were:

- Pre-selected Checkboxes: Options that unknowingly enroll users in services or subscriptions
- Confirm Shaming: Techniques that use guilt or shame to coerce users into opting for something.
- Hidden Costs: Additional, often obscured fees that are not revealed until the final stages of a transaction.
- Fake Urgency: False notifications or countdowns suggesting a limited window of opportunity, pressuring users to make hasty decisions.

Additionally, we included four user interfaces that exemplify persuasive but not deceptive design elements, each accompanied by clarifications to highlight their legitimacy:

- GDPR compliant cookie banner [16]: This banner is a regulatory requirement for obtaining user consent for data collection and processing. The design is straightforward, providing clear information about data usage and allowing users to make informed choices.
- Personalized Recommendations: These recommendations are based on the principle of personalization within persuasive technology [13]. This approach persuades users by providing value and relevance, improving their overall interaction with the platform without any manipulation.
- Price Reductions: Genuine price reductions, as opposed to deceptive practices like inflating prices to create a false discount [17], are examples of persuasive design. When discounts are truthful and clearly communicated, they provide a compelling incentive for users to make purchases.
- Native Advertising: Also known as sponsored content, native advertising is a type of paid advertising that appears in the style and format of the content close to where the advertisement is placed, often seamlessly integrated within a publisher's site or across media. They typically include a headline, an image, and a brief description designed to attract clicks. FTC regulations require that such advertisements be clearly identified as advertisements so that they do not become deceptive practices [18].

We developed eight distinct user interfaces using Figma, a collaborative web application for interface design. These interfaces included both deceptive (dark patterns) and non-deceptive elements to assess participants' ability to distinguish between them. Participants viewed the eight interfaces in a randomized order to prevent any order effects.

We also asked them to rate their perceived manipulativeness of the presented user interfaces, with 1 being the lowest and 5 being the highest, to gauge the varying degrees of certainty in participants' responses. This process aimed to gain insights into how users discern dark patterns from legitimate design practices.

To ensure the integrity of this survey, it is important to note that the interfaces are simulations of real-world website designs, created to reflect the variety of user experiences encountered online. Throughout the survey, participants were informed that the focus was on their perceptions of different user interfaces. However, specific details about dark patterns were withheld until the follow-up survey. This approach was adopted to ensure an unbiased evaluation of the interfaces, allowing participants to form their own opinions without preconceived notions about dark patterns influencing their judgments.

B. Analyses and Results

We incorporated Signal Detection Theory (SDT) into our analysis to elucidate decision-making processes under conditions of uncertainty. SDT, developed by Green and Swets [19], helps distinguish between genuine signals and background noise, allowing us to quantitatively assess participants' ability to identify dark patterns (DPs) within web interfaces. Using SDT, we categorized the responses into Hits, Misses, False Alarms, and Correct Rejections. This categorization helps us understand the accuracy of participants in detecting manipulative designs.

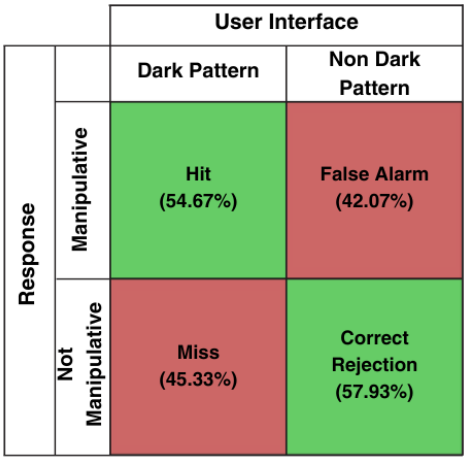

Fig. 1. SDT Analysis of Participant Responses toward User Interfaces.

The horizontal axis represents the type of user interface (Dark Pattern vs. Non-Dark Pattern), while the vertical axis represents the participants' response (Manipulative vs. Not Manipulative). Green sections indicate correct identifications (Hits and Correct Rejections), while red sections indicate incorrect identifications (Misses and False Alarms).

- Hit (54.67%): When participants correctly identified dark patterns, showing they recognized manipulative tactics.
- Miss (45.33%): When participants failed to spot dark patterns, showing how well these designs can blend in with normal ones.
- False Alarm (42.07%): When participants mistakenly thought normal designs were dark patterns, indicating they were cautious, sometimes overly so.
- Correct Rejection (57.93%): When participants correctly identified normal designs as non-deceptive, showing they could tell the difference between deceptive and non-deceptive designs.

To provide a more detailed view of these responses, Table I breaks down participants' perceptions of each interface. This table indicates whether participants classified each interface as manipulative, non-manipulative, or remained unsure:

TABLE I. PARTICIPANTS PERCEPTIONS OF INTERFACE MANIPULATIVENESS

User Interface	Manipulative	Non-Manipulative	Unsure/Don't Know
DP: Pre-selected checkboxes	41.36% (n=146)	46.74% (n=165)	11.90% (n=42)
DP: Confirm shaming	77.05% (n=272)	16.43% (n=58)	6.52% (n=23)
DP: Hidden Cost	35.98% (n=127)	53.82% (n=190)	10.20% (n=36)
DP: Fake Urgency	79.89% (n=282)	15.30% (n=54)	4.82% (n=17)
NON-DP: GDPR compliant cookie banner	35.41% (n=125)	49.29% (n=174)	15.30% (n=54)
NON-DP: Relevant recommendations	46.18% (n=163)	44.48% (n=157)	9.35% (n=33)
NON-DP: Price Reductions	60.91% (n=215)	32.58% (n=115)	6.52% (n=23)
NON-DP: Native Advertising	45.33% (n=160)	44.48% (n=157)	10.20% (n=36)

The information from Table I is also visualized in fig. 2:

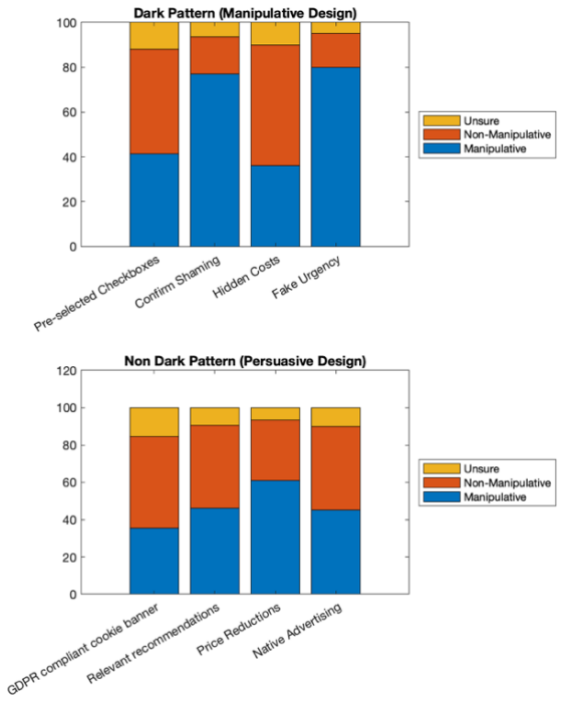

Fig. 2. Perceptions of Manipulativeness in Different User Interfaces

Fake urgency and confirm shaming dark patterns are notably recognized as manipulative. As observed from Figure 2, Fake Urgency, exemplified by sale countdown timers, was recognized as manipulative by a substantial 79.89% (n=282) of participants. Similarly, Confirm Shaming, with phrases such as "No thanks, I don't like saving money," was identified by 77.05% (n=272) of respondents, underscoring its visibility as a deceptive design. Conversely, dark patterns such as Pre-selected Checkboxes and Hidden Costs were less overt, suggesting more subtle manipulation that may not be immediately apparent to users. This nuanced perception of dark patterns highlights the varied impact these designs can have on user trust and behavior.

Furthermore, our results showed an intriguing trend among the control UIs, with a significant proportion of participants perceiving Native Advertising, Personal Recommendations, and Price Reductions as potentially manipulative. This suggests that even non-deceptive design elements can be misconstrued as manipulative, highlighting the complexity of user perception and the importance of clear, transparent communication in design practices.

IV. Discussion

In our study, the distinction between participants' responses to various dark patterns, such as Fake Urgency and Confirm Shaming versus more subtle manipulations, such as Pre-selected Checkboxes and Hidden Costs, provides insightful revelations into user perception.

As mentioned in the literature review on the default effect, when faced with predefined options or default settings, our natural inclination is to adhere to the choices presented. Instinctively, we tend to 'follow the path laid out for us', proceeding without considering alternatives and accepting the choices made on our behalf [11]. That said, more subtle deceptive patterns fit more seamlessly into the user's journey, making them less noticeable.

Conversely, participants in our study perceived some non-deceptive user interfaces, such as Relevant Recommendations and Price Reductions, as manipulative. Below are some possible explanations, along with recommendations for addressing them:

- **User skepticism:** User skepticism is often triggered by the perception that price reductions are "too good to be true," therefore they are manipulative [20]. To address this, for legitimate discount campaigns, businesses could benefit from close collaboration between marketing and design teams. By creating cohesive campaigns and disseminating updates across multiple channels, including social media, companies can effectively inform users and increase the credibility and legitimacy of their offers.
- **Lack of transparency:** When handling user data, it is crucial to provide clear explanations of how personal data is used, including for personalized recommendations and cookies. Transparency includes clearly stating how data is obtained and for what purposes it is used [16]. Implementing a simple hover-over information display after the "you might like this" title could alleviate concerns by offering brief insights into data usage and privacy practices. Example for the hover feature is depicted in Fig.3.

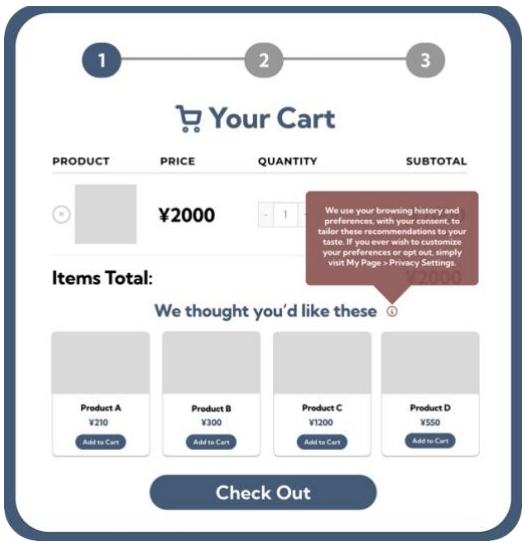

Fig. 3. Example of a Hover-Over Information Feature for Enhanced Data Transparency.

Ultimately, the core of ethical design is about intention and how values are communicated to the user. It is about how we engage users in understanding and navigating digital environments. Design practitioners might focus on user research and usability testing to ensure that design elements effectively convey the intended message.

V. Limitations and Future Research

This study, as indicated in the title, focuses on an exploratory investigation within Japan. While it provides valuable insights into user perceptions, it does limit the generalizability of the findings to other populations. Future research should explore similar questions in different contexts to validate these findings and provide a broader understanding.

One limitation of our study is the presentation of user interfaces in a static format, which may not fully capture the interactive and responsive nature of real-world web environments. Consequently, the applicability of our findings to real-world user experiences may be limited. Future research using interactive website simulations could provide a more realistic context for studying user interactions and yield insights closer to real-world scenarios.

Additionally, a deeper investigation is needed to understand why certain aspects are perceived as manipulative. Blind reviews of design interface manipulativeness, coupled with follow-up interviews, are necessary to explore these complexities further. This could provide valuable insights into the underlying reasons behind user perceptions and help refine design practices.

VI. Conclusion

When used responsibly, persuasive design can enhance the user experience and guide decisions without compromising freedom of choice. However, designers and companies must tread carefully, as even well-intentioned nudges can inadvertently cross into manipulative territory as evidenced by "dark patterns" in UX design.

Our recommendations for UX practitioners, to steer clear of the gray areas that blur the line between persuasion and manipulation could include the following proactive steps: Starting from referring to the established UX guidelines [21],

confirming with the relevant stakeholder guidelines (e.g., European Data Protection Board in the European Union, Federal Trade Commission in the United States, Consumer Affairs Agency (CAA) in Japan, etc.). Most importantly, when conducting usability testing, do not hesitate to ask users about their feelings. Explore potential feelings of difficulty, loss of autonomy, or perceived limitations. Using targeted questions can help uncover underlying issues efficiently.

REFERENCES

[1] R. H. Thaler and C. R. Sunstein, *Nudge: the final edition*. London, UK: Penguin UK, 2012.

[2] D. Susser, B. Roessler, and H. Nissenbaum, "Online Manipulation: Hidden influences in a digital world," *Georgetown Law Technology Review*, vol. 4, no. 1, Dec. 2018, doi: 10.2139/ssrn.3306006.

[3] A. Mathur et al., "Dark patterns at scale: Findings from a crawl of 11K shopping websites," in *Proc. ACM Hum.-Comput. Interact.*, vol. 3, no. CSCW, 2019. doi: 10.1145/3359183.

[4] H. Brignull, "Deceptive Patterns." https://www.deceptive.design/ (accessed Mar. 31, 2024).

[5] H. R. Arkes and P. Ayton, "The sunk cost and concorde effects: Are humans less rational than lower animals?" *Psychological Bulletin*, vol. 125, no. 5, pp. 591–600, Sep. 1999, doi: 10.1037/0033-2909.125.5.591.

[6] C. M. Gray, Y. Kou, B. Battles, J. Hoggatt, and A. L. Toombs, "The dark (patterns) side of UX design," in *Proc. 2018 CHI Conf. Human Factors Comput. Syst.*, Montréal, Canada, 2018, pp. 1-14. doi: 10.1145/3173574.3174108.

[7] "Bringing Dark Patterns to Light," *FTC*. https://www.ftc.gov/reports/bringing-dark-patterns-light (accessed Mar. 31, 2024).

[8] "Guidelines 03/2022 on deceptive design patterns in social media platform interfaces: how to recognize and avoid them," *European Data Protection Board*. https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-032022-deceptive-design-patterns-social-media_en (accessed Mar. 31, 2024).

[9] N. Panjaitan and K. Watanabe, "Deceptive patterns in Japan's digital landscape: Insights from user experience," to be presented at the 26th Int. Conf. Human-Computer Interaction, Washington D.C., USA, 2024.

[10] A. Tversky and D. Kahneman, "Judgment under uncertainty: Heuristics and biases," *Science*, vol. 185, no. 4157, pp. 1124–1131, Sep. 1974, doi: 10.1126/science.185.4157.1124.

[11] E. J. Johnson, S. Bellman, and G. L. Lohse, "Defaults, framing and privacy: why opting in-opting out1," *Marketing Letters*, vol. 13, pp. 5–15, Jan. 2002, doi: 10.1023/A:1015044207315.

[12] L. Mittone and L. Savadori, "The scarcity bias," *Applied Psychology*, vol. 58, no. 3, pp. 453–468, Jun. 2009, doi: 10.1111/j.1464-0597.2009.00401.x.

[13] B. J. Fogg, *Persuasive technology : using computers to change what we think and do*. Burlington, MA, USA: Morgan Kaufmann Publishers, 2003.

[14] A. Mathur, J. Mayer, and M. Kshirsagar, "What makes a dark pattern... dark? Design attributes, normative considerations, and measurement methods," in *Proc. 2021 CHI Conf. Human Factors Comput. Syst.*, New York, NY, USA, 2021, pp. 1-18. doi: 10.1145/3411764.3445610.

[15] L. S. Chamorro, K. Bongard-Blanchy, and V. Koenig, *Ethical Tensions in UX Design Practice: Exploring the Fine Line Between Persuasion and Manipulation in Online Interfaces*. 2023, pp. 2408–2422. doi: 10.1145/3563657.3596013.

[16] "GDPR compliance checklist," *GDPR EU*. https://gdpr.eu/checklist/ (accessed Mar. 31, 2024).

[17] "Deceptive pricing: 16 CFR Part 233," *FTC*. https://www.ftc.gov/legal-library/browse/rules/deceptive-pricing (accessed Mar. 31, 2024).

[18] "Native Advertising: A Guide for Businesses," *FTC*. https://www.ftc.gov/business-guidance/resources/native-advertising-guide-businesses (accessed Mar. 31, 2024).

[19] D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*. John Wiley, 1966.

[20] P. De Pechpeyrou and P. Odou, "Consumer skepticism and promotion effectiveness," *Recherche Et Applications En Marketing (English Edition)*, vol. 27, no. 2, pp. 45–69, Jun. 2012, doi 10.1177/205157071202700203.

[21] "Design principles," *Interaction Design Foundation*. https://www.interaction-design.org/literature/topics/design-principles (accessed Mar. 31, 2024).