DARK PATTERNS

MALICIOUS INTERFACE DESIGN FROM A USERS' PERSPECTIVE

Kristi Bergman
(5701791)

MASTER THESIS

Human Computer Interaction
University of Utrecht
5th July 2021

Examiners:
Christof van Nimwegen
Almila Akdag

Abstract

Dark Patterns are user interface design choices that benefit an online service by coercing, steering, or deceiving users into performing certain actions they did not intend to do. The various malicious strategies that were identified as 'Dark Patterns' were found to be prevalent on numerous online platforms, and were also found to be highly effective. To prevent Dark Patterns from continuing to harm users, awareness of Dark Patterns needs to be created, and a better understanding of Dark Patterns needs to be developed. The current study aims to contribute to this objective by looking at Dark Patterns from a user's perspective. First of all, it investigates the effect of Dark Patterns upon the *user's experience*. Second, it captures a user's take upon the *level of severity* of each of the Dark Pattern types, and turns this into a 'Dark Pattern Darkness Score' (DPDS), which can be used to classify Dark Pattern types based upon their severity-impact. Lastly, the current study takes the first steps into developing a 'System Darkness Scale' (SDS), which can be used to evaluate the *'darkness'* of a system as a whole.

Keywords **–** Dark Patterns, Malicious Interface Design, Deception, Malicious Designer Intent, User Experience, Severity Scores, System Darkness Scale, Dark Pattern Darkness Score.

Contents

1 Introduction

1.1 Scope

Usually, when interaction designers design a product or service, they use general guidelines and methods that all have been widely accepted in the community as good practice. Therefore, when we encounter a poorly designed interface, we often think of it as bad design, which can be the result of laziness or inexperience of the designer (Kitsing, 2018).

There exists, however, a more sinister approach to designing interfaces. Sometimes bad design does not happen accidentally, but rather is created intentionally. Such design is often crafted to confuse users, make it difficult for users to express their actual preferences, or manipulate users into taking certain actions (Luguri and Strahilevitz, 2021). When an interface is maliciously crafted, with a solid understanding of human psychology, to deceive users into performing actions they did not intend to do, we are talking about *Dark Patterns* (Brignull et al., 2015).

Probably the most famous (and very first) example of a Dark Pattern is the advertisement from the Chinese sneaker manufacturer 'Kaiwei Ni' on Instagram Stories. The designer of the advertisement designed the ad in such a way that it looks as if there is a strand of hair on the screen of your phone. The goal of the designer was to trick Instagram users into swiping the hair off their screen. However, as the hair is not physical, but photo shopped *in* the advertisement, Instagram users end up swiping up on the ad, opening the website of 'Kaiwei Ni' by accident (Medium.com, 2017).

Figure 1: Famous Dark Pattern: Instagram Advertisement containing a stray hair. Image retrieved from Medium.com (2017).

These, and many other malicious design strategies have led to various conversations among researchers within the Human Computer Interaction (HCI) community, each of them addressing the (un)ethical nature of Dark Patterns. While (interface) design is by definition a persuasive act and has the potential to manipulate users, the HCI community agrees that in the case of Dark Patterns, designers abuse their power. As such, the HCI community is working hard towards more 'Ethical UI Design' (Di Geronimo et al., 2020). They do this by creating a better understanding of Dark Patterns, by increasing awareness of the deceiving nature of Dark Patterns, and by proposing alternative design solutions.

The growing amount of publishing on Dark Patterns has led to some positive developments within the field of interface design already. However, Dark Patterns still remain an understudied topic in the HCI literature. As such, the current study aims to contribute to creating a better understanding of the various malicious design strategies. More specifically, this study aims to provide an insight into (1) the user's experience (e.g. feeling) when encountering a Dark Pattern, and (2) the severity of each of the various Dark Patterns. Additionally, the current study aims to take the first steps into developing a scale to measure the 'overall darkness' of a specific system.

1.2 Outline

To provide some insight in the current state of research, the next section discusses some key works on Dark Patterns. The concepts, theories and tools that led to the rise of Dark Patterns will be described, different types of Dark Patterns will be explained, and results of existing research on the prevalence and effectiveness of Dark Patterns will be presented. In addition, some cognitive and psychological theories responsible for the functioning of Dark Patterns are discussed.

Based on the insights of the gathering of related work, three clear gaps could be identified. As such, the current study will be threefold, consisting of (a) an experiment on the effect of Dark Patterns on user experience, (b) the assignment of severity (or 'darkness') scores to the different types of Dark Patterns, and (c) the development of a scale to measure the 'darkness' of a service or product as a whole.

Section 3 describes the proposed research methodology for all three parts of the study. The results that were gathered for all three study parts are presented in the Results section and interpreted in the Conclusion section. The Discussion entails the limitations and opportunities of the current study, and provides possibilities for further research.

2 Related work

This section aims to discuss some of the relevant literature in the domain of Dark Patterns. Before we dive deeper into the concept of Dark Patterns itself, the road to the emergence of Dark Patterns will first be discussed. We will start off with the field in which Dark Patterns have emerged, and end with the persuasive technologies that turned into malicious practices later on.

After we have created a better understanding on the background of Dark Patterns, we will focus on these misleading design strategies themselves. Different types of Dark Patterns will be discussed, as well as their effectiveness, prevalence, and underlying psychological mechanisms.

Lastly, based on the findings of the work related to Dark Patterns, the potential gaps in the existing body of literature will be discussed, and the aims of the current study will be presented.

2.1 Human Computer Interaction

Throughout the last decades, technology has advanced substantially and has spread into people's everyday lives (Kim, 2015). As a result, the number of websites and mobile applications available is growing exponentially at high speed (Maier, 2019). People interact with a subset of these websites and mobile applications on a regular basis, as well as with a variety of other technological devices -such as ticket kiosks and payment systems. To make sure these interactions run smoothly, it is important that the interfaces of these devices are designed in a clear and understandable manner, and match the intentions and expectations of the user (Whitenton, 2018).

The research field of Human-Computer Interaction (HCI) provides insights into the ways humans interact with computing devices, by combining knowledge from various disciplines, among which engineering, psychology, ergonomics, and design (Interaction Design Foundation, 2002).

2.1.1 Norman's Action Cycle

Within the HCI literature, there exist various interaction theories that describe the way humans interact with computer systems. One of the most famous interaction theories was proposed by Norman (2013). In his book "The Design of Everyday Things", Norman (2013) presents a conceptual framework called an *Action Cycle*, which interprets user interaction in three components:

- The gulfs of execution and evaluation
- The seven stages of action
- The three levels of processing

Each component provides a different perspective on human-computer interaction, but the components interconnect with each other as well. Figure 2 provides a visual overview of the whole Action Cycle, with all three components present. In what follows, each of the three components will shortly be discussed.

The Gulfs of Execution and Evaluation

The first component of Norman's action cycle consists of two parts to an action: *execution* and *evaluation*. Usually, when a user is planning on using a system, the user has some goal in mind. To accomplish this goal, the user often has to perform certain actions. *Execution* is concerned with a user figuring out how to perform these actions, by discovering the systems' properties.

After the user has executed certain actions (e.g. interacted with the system), *evaluation* is needed. Based on the feedback of the system, the user needs to figure out whether the performed actions have led to accomplishment of the intended goal.

The two parts to an action can also be described into terms of 'gulfs'. As Norman (2013) states in his 'Design of Everyday Things', the *gulf of execution* involves performing a sequence of steps, whereas the *gulf of evaluation* involves perceiving and interpreting the outcome of performing these steps. The two gulfs are situated between the users' goal and the physical state of the world.

The Seven Stages of Action

The second component of the action cycle consists of *seven stages of action*, which -for most part- are situated within the gulfs of execution and evaluation. A user (unconsciously) goes through each of the stages while interacting with a system. The seven stages are:

1. **Goal:** Forming a goal about something that needs to be accomplished.
2. **Plan:** Forming an intention to act.
3. **Specify:** Selecting a sequence of actions that will lead to the goal.
4. **Perform:** Executing the actions.
5. **Perceive:** Perceiving the state of the world resulting from executing the actions.
6. **Interpret:** Interpreting the perception based on expectations.
7. **Compare:** Evaluating whether or not the intended goal was reached.

Three Levels of Processing

The final component of the action cycle consists of three levels of processing, which are cognitive and emotional responses users experience during and after interacting with the system. The three levels of processing can be associated with the seven stages of action, and essentially yield the gulfs of execution and evaluation. The levels are:

- **Visceral**: The subconscious reaction to aesthetics. A user can experience emotions such as calmness or anxiety. For example, when a user is unfamiliar with the design of a particular interface, he will likely experience feelings of anxiety when performing an action and evaluating it.
- **Behavioural**: The subconscious usability of the system. A user's expectations will drive emotions such as hope and fear, and feelings of relief or despair. For example, while specifying the steps for performing a certain action in an interface, a novice user could experience feelings of doubt. But if the user interprets positive results from his actions, he will experience relief.
- **Reflective**: The conscious rationalisation and intellectualisation. A user will compare his expectations with what has actually happened. For example, if a user has successfully completed his activity, he will likely experience feelings of satisfaction.

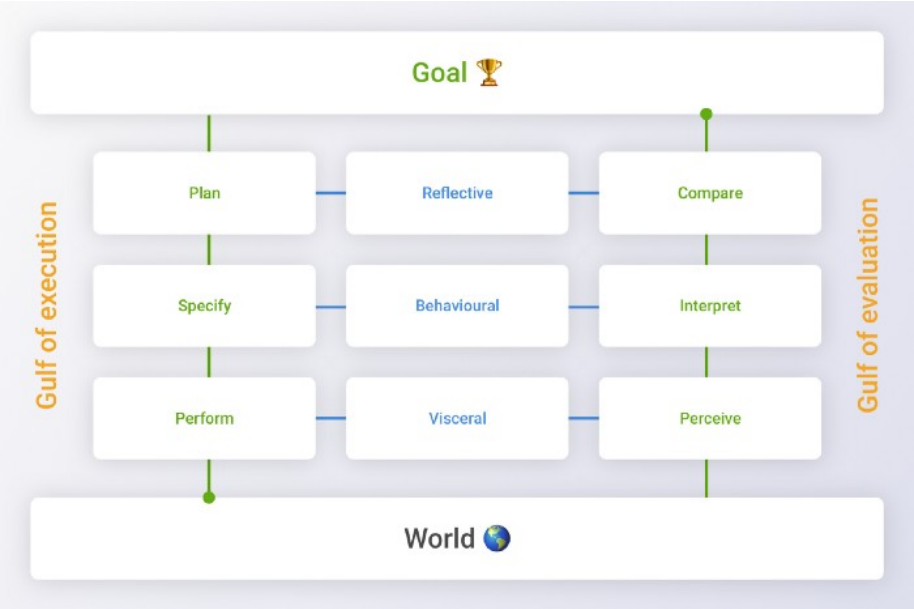


Figure 2: Norman's Action Cycle, with the two gulfs of execution and evaluation (orange), seven stages of action (green) and three levels of processing (blue). Image retrieved from Batterbee (2020).

Practical Relevance

Norman's action cycle provides an insight into the way a user interacts with a system. The action cycle is often used as a guideline to good (interface) design. By being aware of the gulfs of execution and evaluation, with their underlying stages of action and levels of processing, designers can create an interface design that supports users overcome the challenges of understanding the current state of a system and figuring out how to change it in order to reach their goal (Whitenton, 2018). In other words, they can design for a smooth interaction.

2.2 Designing User Interfaces

In simple terms, a *User Interface (UI)* is the features of a device or an application that allow a user to interact with it. A UI communicates to the user where they have arrived, what actions they can perform and how to perform these. The user receives this information from every aspect of the interface design: graphical and textual, silent and audible, static and moving, intentional and accidental (McKay, 2013).

User Interface Design refers to the design of the user interface -it is about programming the look of things (Berezhnoi, 2019). Good user interface design helps users accomplish their desired goals easily and effortlessly. It aids users in reaching their goals by being self-explanatory, presenting information in an understandable fashion, and allowing for easy navigation (Sommerer and Mignonneau, 2008).

2.2.1 Usability

A central concern of UI design is to develop products that are usable. By this what is generally meant is easy to learn and effective to use (Preece et al., 2015). The *usability* of an interface depends on how well its features accommodate users' needs and contexts. When users first encounter an interface, they should be able to find their way through the interface easily enough to achieve their goals without relying on outside- or expert knowledge. A design with high usability guides users through the easiest and least labour-intensive route (Interaction Design Foundation, 2020a). In simpler words: a design with high usability avoids making the user have to think (Láng and Pudane, 2019).

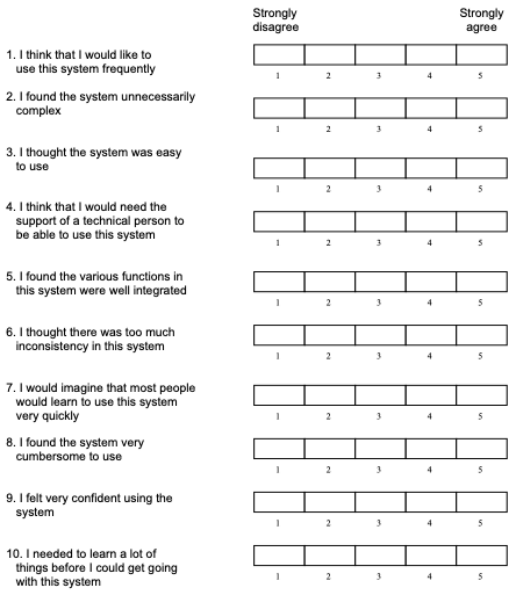


Figure 3: System Usability Scale. Retrieved from Brooke (1996).

2.2.2 System Usability Scale

To measure whether the implemented UI design is indeed perceived as usable, the *System Usability Scale (SUS)* (Brooke, 1996) is often used. It consists of a 10 item questionnaire with five response options, ranging from 'strongly agree' to 'strongly disagree' (figure 3). Although a number of alternatives to the SUS is proposed over the years, the SUS is still deemed as a good choice for general usability practitioners, as it can quickly and easily collect a user's subjective rating of a product's usability (Bangor et al., 2008).

2.2.3 User Experience

Another component that is central to UI design is *User Experience*. Rather than just the interaction, user experience can be seen as the result from interacting with an environment (Buley, 2013). The environment is provided by a product, system or service and the result of interacting with it encompasses the perception, attitude and emotional responses of the user (International Organization for Standardization, 2019). In simpler words, the user experience is about how people *feel* about a product. It includes their overall impression about how pleasant a product is to use (Preece et al., 2015). Products that provide great user experience are therefore designed with not only the product's consumption or use in mind, but also the entire process of acquiring, owning and even troubleshooting it. The focus is not only on creating products that are usable, but also on aspects such as pleasure, efficiency and fun (Interaction Design Foundation, 2020b). It is important to note that one cannot design a user experience, but only design *for* a user experience by creating design features that evoke it (Preece et al., 2015).

2.2.4 (Customer) Journey Maps

To measure whether the implemented design features lead to the desired user experience, designers often make use of so called *(Customer) Journey Maps*. Journey Maps are visualisations of a users' interactions with a product or service (Marquez et al., 2015; Micheaux and Bosio, 2019). They show the user journey from start to finish on completing a task, and map the significant changes in the users' needs, degrees of satisfaction, or other metrics of interest across the various stages, steps and touch points a user must pass through in order to complete the task (Howard, 2014). Typically, the horizontal axis of a journey map represents the progression of time, whereas the vertical axis of the map represents the metric of interest. An example of a journey map is provided in figure 4.

Identifying the Touch Points Customer Journey Maps can be developed in multiple ways. To identify all the touch points a user experiences during performing a certain task (horizontal axis), the *observation technique* is often used. Using this technique, a researcher records all the interactions between a user and a particular service or product (for multiple users). This results in a list of a variety of touch points. As it is important that the customer journey map is representative for the majority of users (e.g. does not include touch points that only very few users experienced), the initial list of touch points needs to

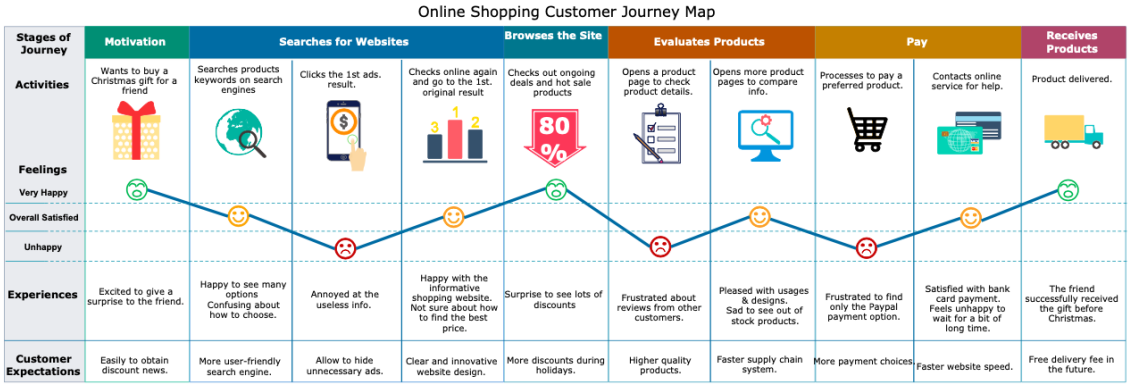


Figure 4: Example of a Customer Journey Map. Retrieved from Edraw (n.d.).

be turned into a final one. To do this, the initial list of touch points can be analysed through absolute frequency. An inclusion criteria of 'a certain percentage of occurrence of a touch point' (for example, 50%) determines which touch points are most important, and therefore are included in the customer journey map (Canfield and Basso, 2017).

Another approach to identify the most important touch points is to let service- or company owners / employees themselves create a list with all possible touch points users might experience during interaction with their product or service. The resulting list of touch points can then be presented to a sample of users, who are asked to indicate -for each touch point- whether they do or do not engage in the touch point when interacting with the product or service (Rosenbaum et al., 2017).

Mapping the Metric of Interest As was already mentioned, the vertical axis of a customer journey map represents the *emotional* journey of users' thoughts, beliefs and feelings that cannot be observed directly. To identify the emotional aspects of a user at each touch point, user research is important. Interviews, (feedback) surveys, social media listening and competitive intelligence are all techniques that provide insights into the user's thoughts and feelings (Interaction Design Foundation, n.d.). Based on the qualitative data that is gathered, companies often develop several personas (each with different personalities, needs and goals), and use these personas to fill out the vertical axis of the Customer Journey Map.

Quantitative data can also provide valuable insights into the user's emotional experience (Chapin, n.d.). For example, a significant drop off in user engagement at a particular screen in a sign-up process could indicate that users did not understand how to proceed and -as a result- got frustrated. Remarkably, however, quantitative application of the Customer Journey Map is scarce.

Goal of Customer Journey Maps The goal of a customer journey map is to aid the design team of a certain product or service in identifying actionable opportunities for improving the user experience (Chapin, n.d.). By mapping the user's emotional experiences at various touch points, designers get an insight into which steps within a certain action are pain points, and therefore which interfaces or design features need re-design or more consideration (Howard, 2014).

2.3 Influencing User Behaviour

Whereas usability once was the key differentiator when it comes to designing effective user interfaces, designing interfaces that are simply easy to navigate and understand are no longer enough today (Schaffer, 2009). Although applying usability techniques will enable users to perform an action efficiently, this does not mean that this action will actually take place. In other words: Just because users *can* do something does not guarantee that they *will*. In order to achieve the latter, users must be motivated and persuaded (Gubaidulin, 2016). As such, design became inherently a persuasive act (Nodder, 2013).

2.3.1 Persuasive Design

The field of *Persuasive Technology* aims to influence the behaviour of people by the use of design. Fogg introduced the term *persuasive technology* in 2003, defining it as "a computing system, device, or application intentionally designed to change a person's attitude or behaviour in a predetermined way". This shaping of behaviour is accomplished with the help of behavioural insights from psychology, such as the existence and operation of human cognitive biases.

Table 1: Seven Persuasive Design Strategies as proposed by Fogg (2003).

Strategy	**Description**
1. Reduction	Simplifies a task that the user is trying to do, usually by eliminating some of the steps necessary in the step sequence to achieve a certain goal. This can lead to the user being encouraged to perform the task and to the user believing in his abilities, tackling the task with a positive approach.
2. Tunnelling	Guides the user towards his goal through a sequence of activities, step by step.
3. Suggestion	Gives suggestions to the user at the right moment and in the right context.
4. Tailoring	Provides a personalised experience to the user in terms of the presented interface, information, options and feedback, based on the user's needs and actions.
5. Self-monitoring	Enables the user to track his own behaviour, in order to inform the user about how they might modify their behaviour to achieve a predetermined outcome.
6. Conditioning	Relies on providing reinforcement (or punishments) to the user.
7. Surveillance	Observes the user overtly in order to increase the intended target behaviour, as when people know they are being watched, they will try to make their actions meet the observer's expectations.

Fogg (2003) listed seven persuasive tools or strategies to influence attitudes or behaviour. The strategies are summarised in table 1. Often, the strategies are used together as part of a system to create a persuasive experience. They can easily be transferred for usage in a (UI) design context.

By applying psychology insights to interface design, designers are able to communicate information to users more precisely, aid users in making decisions, nudge users toward completion of their goal, assist them in developing new skills, and even help users end or begin new habits (UX Booth, 2018).

Studies of persuasive technology have shown the potential benefits of designing for persuasion. One good example of such can be found within the health domain, where persuasive technologies were deployed to promote physical activity (Consolvo, Everitt et al., 2006; Consolvo, Klasnja et al., 2009; Lin et al., 2006). Other applications of persuasive technology within this domain include technologies which motivate users to quit smoking (Paay et al., 2015), lose weight (Arteaga et al., 2009; Peng, 2009) or sustain a healthy water intake (Chiu et al., 2009).

2.3.2 Digital Nudging

Another concept that one often encounters when it comes to influencing (user) behaviour is *nudging*. Digital nudging can briefly be described as "the use of user interface design elements to guide people's behaviour in a digital environment, without restricting the individual's freedom of choice" (Meske and Potthoff, 2017). This latter part of the definition of nudging is important, as the main intention of nudging is to increase people's long-run welfare, by helping them make better choices, without forcing certain outcomes (Thaler and Sunstein, 2009).

Table 2: Summary of Nudge principles. Retrieved from Weinmann et al. (2016).

Nudge Principle	**Description**	**Example**
i**N**centive	Making incentives more salient to increase their effectiveness.	Telephones that are programmed to display the running costs of phone calls.
Understanding mappings	Mapping information that is difficult to evaluate to familiar evaluation schemes.	Mapping megapixels to maximum printable size when advertising a digital camera instead of pointing to megapixels.
Defaults	Preselecting options by setting default options.	Automatic renewal of subscriptions.
Giving feedback	Providing users with feedback when they are doing well and when they are making mistakes.	Electronic road signs with smiling or sad faces depending on the drivers speed.
Expecting error	Expecting users to make errors & being as forgiving as possible.	At an ATM, people must retrieve their card before they receive their money in order to help them avoid forgetting their card.
Structure complex choices	Listing all the attributes of all the alternatives & letting people make tradeoffs when necessary.	Online product configuration systems that make choices simpler by guiding users through the purchase process.

Weinmann et al. (2016) summarised the 'principles' used in digital nudging, and presented a nudging example for each of these principles. The summary is presented in table 2.

2.4 From Persuasion to Deception

While persuasive technology and nudging practices are often praised for the good they are capable of producing in society and individual life, there are also substantial ethical considerations regarding designing explicitly to persuade (Gray et al., 2018). Reason for this might be that, when designing user interfaces, there is a thin line that stands between persuasion and deception (Láng and Pudane, 2019).

To distinguish the two, we can look at the *intentions* of the designers. In persuasive and nudging practices, designers aim to motivate or encourage users to freely explore specific content and take certain action. These practices are generally designed to guide user's behaviour with the goal to make users better off. To reach this goal, persuasive practices use techniques in which the user is put at the centre of attention (Láng and Pudane, 2019; Sunstein, 2018).

In deceptive practices, on the other had, designers aim to either trick the user into taking an action or prevent the user from performing it. In this case, the persuasive techniques are no longer deployed with the user in mind, but are rather business centric (Gray et al., 2018; Nagda, 2020).

2.4.1 Unintentional Design Failures

Sometimes deception occurs unintentionally. Due to a lack of technical skills, inexperience or little knowledge of user needs, a designer can design a non-working solution that results in an unintended negative user experience (Greenberg et al., 2014). Such a design solution is often called an *anti-pattern*. When an anti-pattern is discovered, it is often documented as 'known bad practice', so use of the design solution can be prevented in future UI design.

2.4.2 The Rise of 'Dark Patterns'

When deception occurs on purpose, instead of unintentionally, we deal with the so called *Dark Patterns* (De Rosis et al., 2003). The term *Dark Patterns* was coined by Harry Brignull in 2010, who defined it as "user interfaces that have been carefully crafted with an understanding of human psychology to trick users into doing things they did not intend to do". In other words, Dark Patterns are user interface design choices that benefit an online service by coercing, steering, or deceiving users into making decisions that, if fully informed and capable of selecting alternatives, they might otherwise not make (Mathur et al., 2019). Dark Patterns therefore do not have the users interest in mind (Gray et al., 2018). At best, Dark Patterns annoy and frustrate users. At worst, they can mislead and deceive users (Mathur et al., 2019).

2.4.3 Digital Nudges, Anti-Patterns & Dark Patterns

Figure 5 presents the differences and similarities of Digital Nudges, Persuasive Technologies, Anti-Patterns and Dark Patterns by means of the two characteristics that are most important in distinguishing these design solutions. Whereas the vertical axis of the graph represents whether the outcome of the design solution was either intended or not, the horizontal axis represents the outcome itself (e.g. the result of the user interacting with the design). The colour of each of the design solutions corresponds to the design focus, which could be either user-centric (green) or business-centric (red).

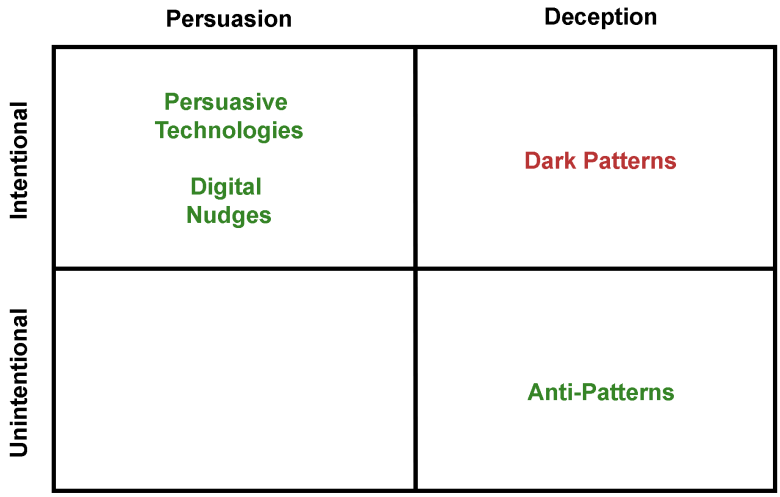


Figure 5: Distinguishing the various Design Solutions.

2.5 Motivations to use Dark Patterns

As mentioned in the previous section, Dark Patterns are user interface design choices that *benefit an online service*. Research has found that dark patterns perform well in multivariate tests and A/B testing (Brignull, 2011; Keith, 2017). As such, applying them will likely result in more sales, higher revenues and the obtainment of more (personal) user data in comparison to an interface design that does not intentionally trick the human mind. This makes dark patterns a valuable and highly effective asset in trying to reach business-oriented goals (Maier, 2019).

The effectiveness of Dark Patterns -resulting in businesses using them- will be discussed in more detail in section 2.7. Some different types of Dark Patterns will be discussed first, in order to provide an insight into what Dark Patterns might look like.

2.6 Dark Pattern Taxonomies

2.6.1 Brignull's Taxonomy

Since the identification of Dark Patterns as an ethical phenomenon in 2010, Brignull has collected a set of artifacs containing Dark Patterns. The artifacts were gathered from blogs, websites, and social media platforms and bundled into an overview, called the "Hall of Shame", on Brignull's website www.darkpatterns.org (Gunnarsson, 2020). The set of artifacts guided Brignull's Dark Patterns Taxonomy, which consisted of 12 types of Dark Patterns, specifically "Bait and Switch", "Disguised Ad", "Forced Continuity", "Friend Spam", "Hidden Costs", "Misdirection", "Price Comparison Prevention", "Privacy Zuckering", "Roach Motel", "Sneak into Basket", and "Trick Questions". All these types of Dark Patterns will be explained in the following subsection.

2.6.2 Gray's Taxonomy

Gray et al. (2018) built upon the the existing taxonomy of Dark Patterns as proposed by Brignull et al. (2015). Whereas Brignull's Taxonomy mixed context, strategy, and outcome, Gray's taxonomy of Dark Patterns is solely based on the strategic motivator behind Dark Patterns (Nagda, 2020). By doing this, the categorisation of Dark Patterns was made sharper, more general, and more suitable for comparison among patterns (Di Geronimo et al., 2020; Gray et al., 2018). As such, Gray's Taxonomy has become the new standard of reference for the UX design community.

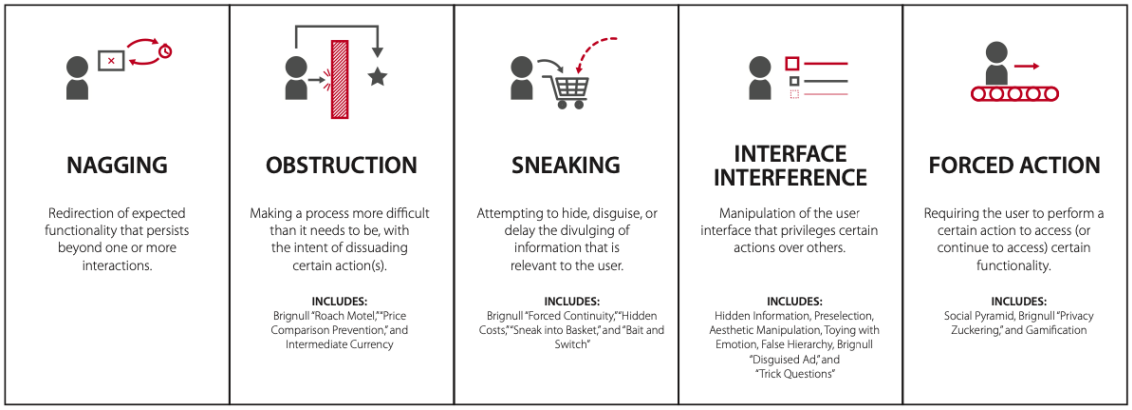


Figure 6: Visual overview of Gray's Dark Pattern Taxonomy. The image shows the five primary Dark Pattern categories and provides a short textual overview of their underlying strategies. Image retrieved from Gray et al. (2018).

Gray et al. (2018) identified five primary Dark Patterns categories that appear to serve as strategic motivators for designers: (1) nagging, (2) obstruction, (3) sneaking, (4) interface interference, and (5) forced action. Each of the five Dark Pattern categories includes multiple Dark Pattern strategies, among which

the original Dark Pattern types as found by Brignull et al. (2015). A visual overview of Gray's Taxonomy is provided in figure 6. In what follows, each of the Dark Pattern categories and their underlying strategies will shortly be discussed. The examples cited within this subsection are all derived from the website https://www.darkpatterns.uxp2.com.

Nagging Gray et al. (2018) defined *nagging* as "a minor redirection of expected functionality that may persist over one or more interactions". Nagging often manifests as a repeated intrusion during normal interaction, where the user's desired task is interrupted one or more times by other tasks not directly related to the one the user is focusing on. Nagging behaviours may include pop-ups that obscure the interface, audio notices that distract the user, or other actions that obstruct or redirect the user's focus.

Examples of nagging can be found within many social media applications, mainly in the form of *persistent pop-ups*. The photo sharing application Instagram (figure 7) prompts users with a pop-up which includes the question to turn on notifications for the app. Only two response options are present: "Not now" and "OK". There is no "No" option for the scenario in which a user does not want to turn on notifications. Therefore, the user is provided with no ability to discontinue the prompts, unless the (by Instagram) desired option of "OK" is chosen.

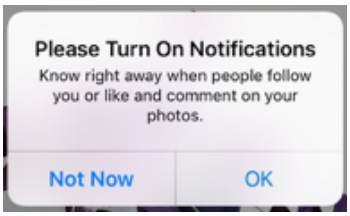


Figure 7: Nagging behaviour on Instagram.

Obstruction A second Dark Pattern category is *obstruction*. Gray et al. (2018) defined obstruction as "impeding a task flow, making an interaction more difficult than it inherently needs to be with the intent to dissuade an action". In other words, obstruction makes a certain process difficult for the user, which makes completion of the process less likely. Obstruction often manifests as a major barrier to a particular task that the user may want to accomplish. There are three Dark Pattern strategies underlying the category of obstruction. Each of them will shortly be explained.

Brignull's Roach Motel

A Roach Motel occurs when a situation is easy to get into, but difficult to get out of. A user is able to sign up for a particular service (e.g., creating an account or subscribing to a mailing list) easily, but cancelling the service is difficult or impossible (Brignull et al., 2015). An analogy for the Roach Motel can be considered the classic maze: Entering a maze is easy, but finding the path out is much more difficult.

A typical example of the Roach Motel Dark Pattern requires a user to call a phone number in order to cancel a subscription. In this case, cancelling the subscription can be made difficult for the user by (1) requiring the user to call during specific (business) hours, and (2) obligating the user to speak to an employee of the subscription-firm, who will pressure the user into maintaining the subscription.

Brignull's Price Comparison Prevention

Humans are known to always seek the better deal. In order to do this, (price) comparisons need to be made. Some online retailers are not confident in letting users freely and easily do price comparisons, as they are afraid that their products will not end up being selected at all or price comparisons will lead to lower profit margins. As a result, the Price Comparison Prevention Dark Pattern is often used.

Price Comparison Prevention seeks to dampen the effect of market competition by making direct price comparisons between products and services difficult. Tactics could include making important product information un-copyable (e.g., product ID, price), in order to prevent users from pasting such information into a search bar or onto another site, or obstructing a direct price comparison between two similar items on the same platform, with the hope of users buying the more expensive option.

An example of the latter tactic can be found on websites of many supermarkets. Figure 8 shows part of a product page of a supermarket in the UK. By providing a user with no opportunity to compare the prices of apples per item or kg for (a) a loose apple and (b) a package of apples, it is difficult for a user to make an informed decision about whether buying six loose apples or a package of six apples is the better (or cheaper) option.

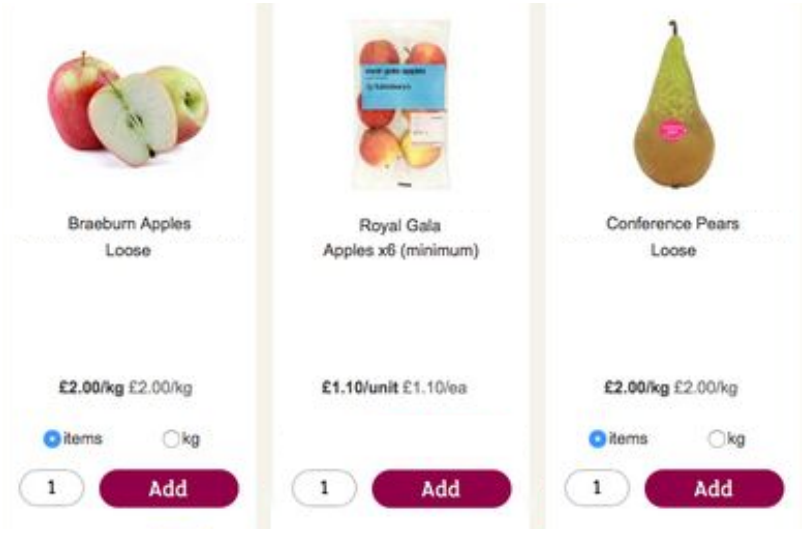


Figure 8: Price Comparison Prevention for apples on a UK supermarket website.

Intermediate Currency

Another Dark Pattern strategy underlying Obstruction is Intermediate Currency. Within this sub type of Obstruction, users spend real money to purchase a virtual currency which is then spent on a good or service. The goal of this pattern is to disconnect users from the real value of money in order to cause the user to interact differently with the virtual currency.

Sneaking A third Dark Pattern category is *sneaking*. Gray et al. (2018) defined sneaking as "an attempt to hide, disguise, or delay the divulging of information that has relevance to the user". Sneaking often occurs in order to make the user perform an action they may object to if they had knowledge of it. Sneaking behaviours may include additional undisclosed costs or undesired effects from a particular action. There are four Dark Pattern strategies that involve sneaking:

Brignull's Forced Continuity

The Forced Continuity pattern continues to charge the user after the service they have purchased expires. The pattern takes advantage of users' failure to check upon service expiration dates, either for a free trial or for limited-time use of a paid service. As a user does not remember to opt out of the service, it is assumed that the user either wants to continue the service, or upgrade to the paid version of the free trial. As a result, the credit card of the user silently and automatically continues to be- or starts getting charged without any warning (Brignull et al., 2015).

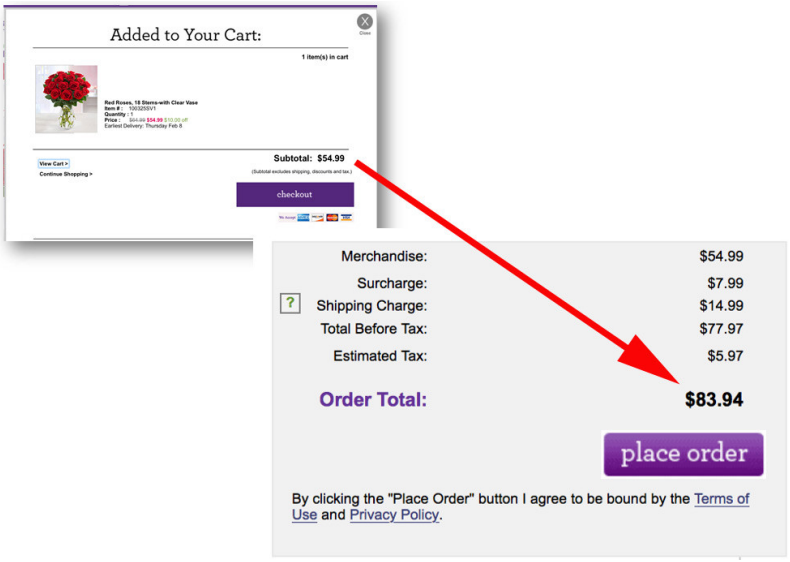


Figure 9: Hidden Costs on the 1-800-Flowers website.

Brignull's Hidden Costs

The Hidden Costs Dark Pattern first shows users a certain price for a service or product, only to reveal new, additional, and often unusually high charges to users later on (Gray et al., 2018). Examples of such charges include service fees, shipping costs, or limited time conditions. Often these charges are disclosed at the end of a checkout process, after the user has already filled out shipping and billing information, and consented to the terms of use (Mathur et al., 2019).

An example of Hidden Costs can be found on the 1-800-Flowers website (figure 9). When a user adds a bouquet of flowers to his shopping cart on this website, the price of this product is displayed as $54.99. After going through the multi-step checkout process, and finally getting to the last step of confirming the purchase, the user ends up with an amount of $83.94 to pay, due to several additional costs, which were not disclosed before (Sauro, 2018).

Brignull's Sneak into Basket

The Sneak into Basket Dark Pattern automatically adds additional products to users' shopping charts without their consent or knowledge. Services who use this type of Dark Pattern often claim that these "sneaked items" are suggestions based on other purchased items, and therefore help the user in finding what they (might) need. However, if the user does not notice the sneaked items prior to checkout, the user might unintentionally buy these items.

One instance of the Sneak into Basket pattern is shown in figure 10. The web shop Sports Direct used to automatically add a Value Magazine and a 'free' mug to all of their online orders for an additional $1. The item could only be removed from the purchase list by deleting the item manually in 'my basket'.

Figure 10: Sneak Into Basket on Sports Direct.

Brignull's Bait and Switch

The Bait and Switch Dark Pattern occurs when a user sets out to do one thing, but a different, undesirable thing happens instead (Brignull et al., 2015). In other words, the action a user performs has another outcome than the user expected (Di Geronimo et al., 2020).

The most famous example of Bait and Switch is Microsoft's approach to getting users to upgrade their Windows OS to Windows 10 (figure 11). Microsoft prompted their users with a pop-up window stating that an update was needed. In most pop-up windows, a user can press the X at the top right corner to close the window, thereby discarding the message. However, in this Windows dialog, clicking the X resulted in the upgrade being initialised -a completely unforeseen result for the majority of the users.

Figure 11: Microsoft's approach to let users upgrade their OS: Bait and Switch.

Interface Interference A fourth Dark Pattern category is *Interface Interference.* Gray et al. (2018) defined Interface Interference as "any manipulation of the user interface that privileges specific actions over others, thereby confusing the user or limiting discoverability of important action possibilities." Interface interference manifests as numerous individual visual and interactive deceptions. The Dark Pattern consists of three sub types:

Hidden Information

The first sub type is Hidden Information. Gray et al. (2018) defined hidden information as "options or actions relevant to the user but not made immediately or readily accessible". To acquire all -and often even critical- information, a user has to put effort in discovering this information.

Hidden information may manifest as options or content hidden in fine print (figure 12a), discolored text (figure 12b), or in a product's terms and conditions statement. In figure 12a, a user must select "More info" to be able to retrieve information about sharing personal data. Important information is 'hidden' in a drop down. In figure 12b, the "unsubscribe here" link is hidden as white text on a white background.

Preselection

A second subtype of Interface Interference is Preselection. Gray et al. (2018) defined preselection as "any situation where an option is selected by default prior to user interaction". Preselection usually manifests as a default choice that the shareholder of the product wishes the user to choose. However, this choice is often against the user's interests or may provide unintended consequences. In other words, in this Dark Pattern, unfavorable options (from the viewpoint of the user) are often preselected, with the aim of the user agreeing to- or overlooking this default option (Di Geronimo et al., 2020).

In figure 12a, an example of Preselection is provided. In addition to information being hidden in a drop down, the information about sharing personal data is also automatically agreed to by the preselection of a checkbox. This example clearly shows the dangers and deceptiveness of Dark Patterns -as combining them can make them even more powerful.

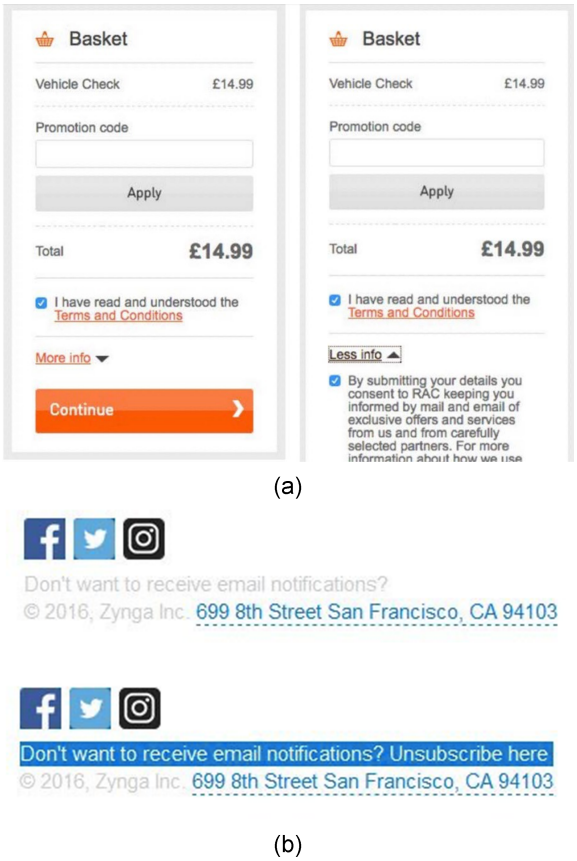


Figure 12: Hidden information in (a) fine print, and (b) discolored text. (a) also shows the pattern of preselection.

Aesthetic Manipulation

A third Dark Pattern strategy underlying Interface Interference is Aesthetic Manipulation. Gray et al. (2018) defined Aesthetic Manipulation as "any manipulation of the user interface that deals more directly with form than function". This includes design choices that focus the user's attention on one thing to distract them from or convince them of something else (e.g., Brignull's "Misdirection"). Four more specific instantiations of Aesthetic Manipulation exist:

Toying with Emotions (Brignull's Confirmshaming)

The Toying with Emotions strategy includes any use of language, style, colour, or other similar elements to evoke an emotion in order to persuade the user into a particular action (Gray et al., 2018). The elements used to evoke emotions appeal to the underlying psychology of the human nature, and aim to elicit a particular response from users (desired by the shareholder) -often through shame and guilt.

An example of Toying with Emotions is provided in figure 13. The website Delish.com prompts the user to sign up for a newsletter. The option for declining is "No thanks, I'll have a microwave dinner tonight". This option could evoke emotions in the user to encourage them to change their intended action of declining to sign up for the newsletter in the first place.

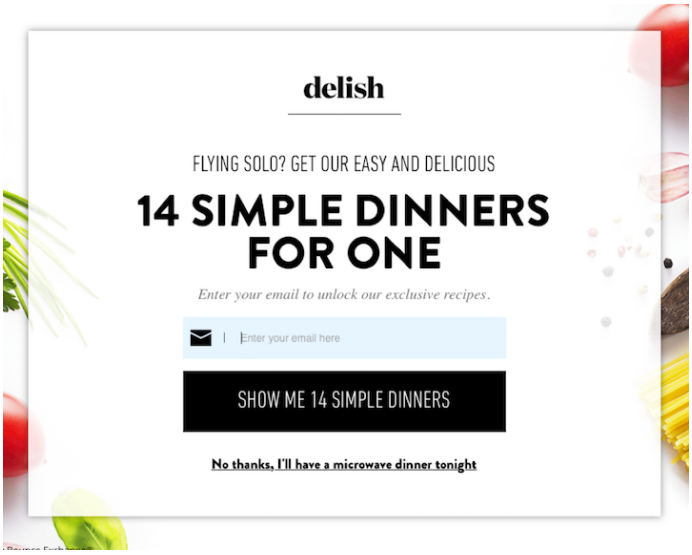


Figure 13: Toying with Emotions on the Delish website.

False Hierarchy

The False Hierarchy strategy gives one or more options visual or interactive precedence over others, particularly where items should be in parallel rather than hierarchical (Gray et al., 2018). This convinces the user to make a selection that they feel is either the only or the best option.

An example of the False Hierarchy pattern is provided in figure 14. Whenever a user wants to install the TuneUp app, the user is asked to make a decision between "express" installation or "custom" installation. The "custom" option is greyed out, giving the user the false impression that this option is disabled. In addition, the "express" installation is given visual precedence by displaying it as "(recommended)".

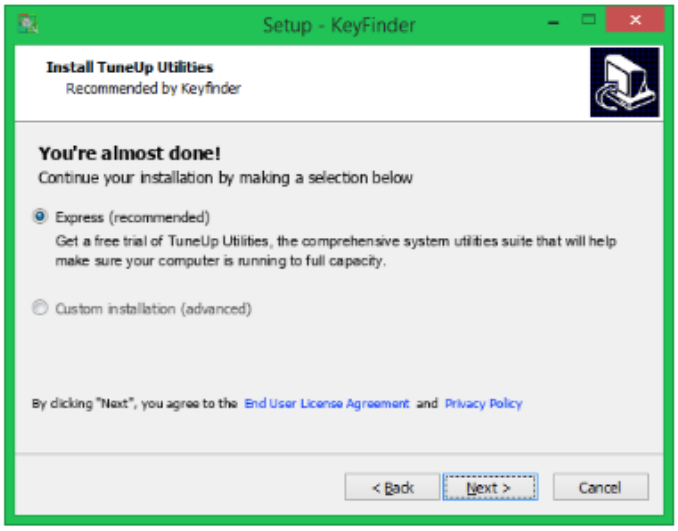


Figure 14: False Hierarchy during the installation process of TuneUp app.

Brignull's Disguised Ad

The purpose of Disguised Ads is to blend with the website, app, or any other medium of information, in order to match the form and function of the platform in which it appears. This way, it is hard for the user to distinguish ads from the natural and organic content of the platform. The ads are often disguised as interactive games, download buttons, or other salient interaction the user is looking for (Gray et al., 2018). By clicking the ad, the user will not perform his intended action, but will be redirected to the ads web page or start an unwanted action.

Figure 15: Ad disguised as interactive game (left), and download button (right).

The ad in figure 15 (left) appears to the user as a game, and uses flash to make it interactive. When a user moves his cursor around the ad, the target moves with it. When the user clicks the ad to 'stomp' the spider, the spider is not actually killed (as would probably happen in a real game), but instead the user is taken to the ads website.

The disguised ad in figure 15 (right) stating "Start Download" (presented right next to the display of letters) really blends in with the website of DaFont. The main download button for downloading The Blacklist font (upper right corner) is much smaller and less visible.

Brignull's Trick Questions

The Trick Questions strategy includes asking a question that appears to be one thing but is actually another, or uses confusing wording, double negatives, or leading language to manipulate user interactions (Gray et al., 2018).

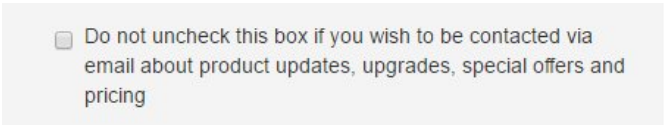


Figure 16: Trick question on codemasters.com using confusing double negatives.

One common example of this tactic is the use of checkboxes to opt out rather than opt in, often paired with confusing double negatives. On the website of Code Masters a double negative is used to trick the user into signing up for newsletters (figure 16).

Forced Action The last category of Dark Patterns is Forced Action. Gray et al. (2018) defined Forced Action as "any situation in which users are required to perform a specific action to access (or continue to access) specific functionality. This action may manifest as a required step to complete a process, or may appear disguised as an option that the user will greatly benefit from."

Social Pyramid

The Social Pyramid strategy requires users to recruit other users to use a particular service. This is a method commonly used in social media applications and online games. Users can invite their friends to use the service and are incentivized with benefits from the platform in return. This pattern subsumes Brignull's "Friend Spam" and expands the definition to include any social recruiting (Gray et al., 2018).

An example of the Social Pyramid strategy is the game FarmVille, in which users are pressured to invite friends. Not specifically because the game is fun, but rather because certain features are useless or inaccessible without online friends also playing.

Brignull's Privacy Zuckering

The Privacy Zuckering Dark Pattern tricks users into sharing more information about themselves than they intend to or would agree to. This includes the selling of user's information to third parties that is included in the Term and Conditions or Privacy Policies of websites (Gray et al., 2018).

Gamification

The Gamification pattern describes situations in which certain aspects of a service can only be "earned" through repeated (and perhaps undesired) use of aspects of a particular service (Gray et al., 2018).

An example of gamification as a Dark Pattern are the impossible levels in the game Candy Crush Saga. The game occasionally gives players levels that are impossible to complete in order to urge them to buy power-ups or extra lives.

Table 3: Summary of all Dark Pattern categories, their underlying strategies, and some of their instances.

Category	**Dark Pattern Strategy**	**Description**	**Examples**
Nagging		Repeated interruption of the users' desired task.	Persistent pop-ups, Audio notices
Obstruction	Roach Motel	Situation is easy to get into, but difficult to get out of.	Cancelling a service or account is impossible
	Price Comparison Prevention	Makes comparing prices between products difficult.	Un-copyable product info, Prices displayed in different units.
	Intermediate Currency	Disconnects users from the real value of money.	In-game credit
Sneaking	Forced Continuity	Automatically charges the user after the service they have purchased expires.	"Free" trial service
	Hidden Costs	Shows users a certain price, but adds additional costs to it in a later stadium of the process.	Unexpected, extremely high fees / costs
	Sneak into Basket	Adds additional items to users' shopping carts without their knowledge or consent.	
	Bait & Switch	Certain action has a different outcome than expected.	

Interface Interference	Hidden Information	Options or actions relevant to the user are not made immediately accessible to them.	Info hidden in fine print, as discolored text, or behind a drop-down
	Preselection	Situation where an option is selected by default prior to user interaction.	Selected checkbox
	Toying with Emotions	Uses language / style / colour to evoke a particular emotion.	Countdown timer, Emotional wording, High-demand messages
	False Hierarchy	Gives one or more options visual precedence over others.	Use of (recommended), Grayed out options
	Disguised Ads	Ads that blend with the medium of information.	Interactive game ads, Fake "download" buttons
	Trick Questions	Question that appears to be one thing, but is actually another.	Confusing wording, Double negatives
Forced Action	Social Pyramid	Requires users to recruit other users to use a particular service.	
	Privacy Zuckering	Tricks users into sharing more information about themselves than they would otherwise agree to.	
	Gamification	Certain aspects of a service can only be earned through repeated use or aspects of the service.	Impossible game levels

2.6.3 Resemblance of Dark- & Persuasive Strategies

Many of the Dark strategies bear striking resemblance to the Persuasive Strategies proposed by Fogg (2003). Persuasive strategies such as 'Tunneling' and 'Reduction' have strong similarity to the 'Forced Action' and 'Obstruction' Dark Pattern strategies. Similarly, 'Tailoring' and 'Suggestion' are akin to strategies underlying 'Interface Interference'. Finally, 'Conditioning' is comparable to strategies such as 'Sneaking' or 'Nagging'. What this indicates is that many persuasive strategies have already been twisted for nefarious purposes (Gray et al., 2018).

2.7 Existing Research on Dark Patterns

As could be seen in the previous section, a great variety of Dark Patterns exist. Despite their 'darkness' and their questionable ethical nature, the prevalence of Dark Patterns on websites, in apps and other (online) platforms is widespread. Therefore, Dark Patterns (indeed) seem to be very effective mechanisms that are able to steer a user into performing an (unwanted) action that benefits the online service.

2.7.1 Prevalence of Dark Patterns

A study by Mathur et al. (2019) used a web crawler to identify Dark Patterns on the 11K most popular shopping websites worldwide. The researchers discovered 1818 instances of Dark Patterns on the shopping websites, which were present on 1254 of them (11.1%). The shopping websites that were more popular, were more likely to feature the Dark Patterns.

A study by Di Geronimo et al. (2020) analysed the prevalence of Dark Patterns in 240 applications. 30 applications were selected for each of the 8 main categories of applications on the Google Play Store. Among the 240 studied apps, Di Geronimo et al. (2020) found that 95% included one of more Dark Patterns in their interfaces. Overall, 1787 Dark Patterns were found among all apps.

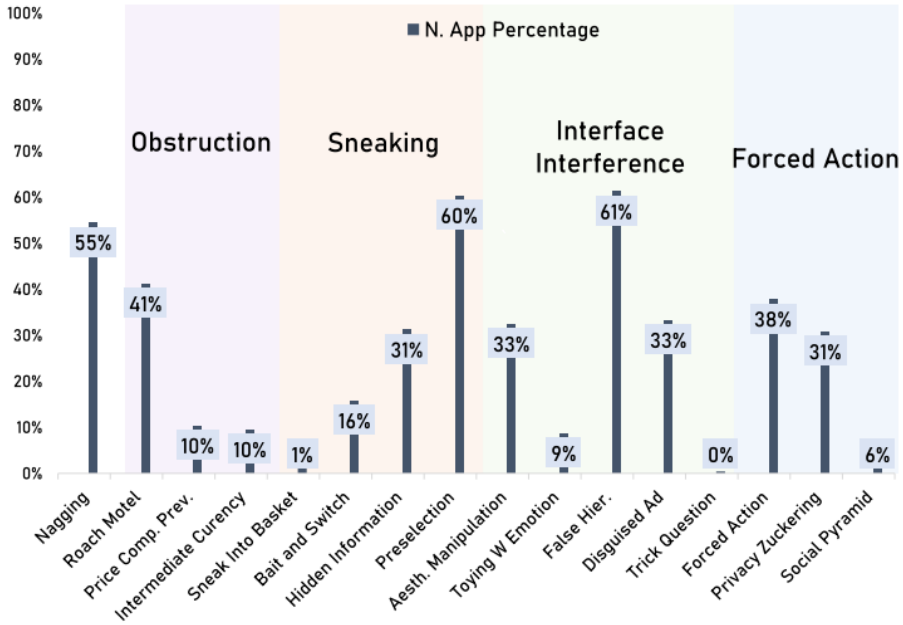


Figure 17: Results of study by Di Geronimo et al. (2020): Percentage of a total of 228 apps containing each of the Dark Pattern strategies as proposed by Gray et al. (2018).

Almost 10% of the apps included 0-2 Dark Patterns, 35% of the apps contained 3-6 Dark Patterns, while the remaining 49% included >7 Dark Patterns. The instances of Dark Patterns found were classified into the taxonomy proposed by Gray et al. (2018). Figure 17 shows the percentage of apps containing each Dark Pattern type. The 12 apps that contained no instances of Dark Patterns were not included here.

2.7.2 Effectiveness of Dark Patterns

The high prevalence of Dark Patterns is not surprising, as numerous studies have found that Dark Patterns are very effective. Most studies on the effectiveness of Dark Patterns have been performed for pop-up or cookie consent requests. Within these studies, user acceptance rates (e.g., sharing of personal information) were used as a measure of effectiveness: The higher the acceptance rate, the more effective the Dark Pattern. In what follows, we will discuss the results of three of these studies.

In a study by Luguri and Strahilevitz (2021), a Bait and Switch scenario was designed, in which participants were asked to fill out a survey that evaluated their attitudes towards privacy. At the end of the survey, participants were deceived into believing that because they expressed a strong interest in privacy, they had been signed up for a costly identity-theft protection service. Participants did have the opportunity to opt out on the service, but the steps that were required to do so varied by the level of Dark Pattern manipulation. Luguri and Strahilevitz (2021) researched three of these levels: (1) no Dark Patterns, (2) mild Dark Patterns, and (3) aggressive Dark Patterns. In the mild Dark Patterns condition, False Hierarchy and Toying with Emotions strategies were used. In the aggressive patterns condition, the Roach Motel, Trick Questions and Nagging strategies were added to this.

The results of the study offered striking empirical support for the proposition that Dark Patterns are effective in bending consumers' will. In the 'no Dark Patterns' condition, only 11.3% of the participants opted to accept the identity protection program. When mild Dark Pattern tactics were deployed, the acceptance rate more than doubled. Now 25.8% of participants accepted the data protection program. When participants were exposed to aggressive Dark Patterns, the acceptance rate went up further, with 41.9% of the sample accepting the program.

In a study by Utz et al. (2019), several different consent pop-up designs were created to investigate the effect of various different design properties on acceptance rates. Amongst the design properties were two Dark Pattern types: Preselection and False Hierarchy.

The experiment revealed that, when given the options 'accept' and 'decline', users are more likely to share personal information when the 'accept' button is given visual precedence over the 'decline' button, as opposed to when the 'accept' and 'decline' buttons were given equal precedence. In addition, the research found that when users are being presented with several checkboxes representing categories that need consent for using personal information (e.g., 'analytics', 'social media' and 'marketing'), the acceptance rate increased from a mere 0.16% to 83.55% when these checkboxes were preselected.

A study by Nouwens et al. (2020) also evaluated the effects of different designs of consent banners on users' consent choices. The study showed that the probability of a user accepting a privacy notice increases with 22 percentage points when the 'reject all' button is removed from the first page of a consent banner (and is hidden on a second page), while the 'accept all' button remains present.

2.7.3 Dark Pattern Awareness / Blindness

When looking at the studies about the effectiveness of Dark Patterns, one might think that users are unconsciously influenced by their appearance, as the exploitation of Dark Patterns leads to users performing different actions as opposed to no exploitation of these malicious patterns. As such, the question arose whether users are indeed unaware of- or actually can recognise Dark Patterns.

In an attempt to answer this question, Di Geronimo et al. (2020) carried out an online experiment in which participants were asked to rate the user interfaces of several applications. Some of these applications included Dark Patterns in their UI, whereas others did not. Participants were presented with videos upon the usage of three applications (two containing Dark Patterns, one without Dark Patterns), after which they were asked whether they spotted any malicious design in the videos.

Results of the experiment showed that the majority of users did not spot malicious designs in the apps containing Dark Patterns (55%). Some were unsure (20%), and the remaining found a malicious design in the apps (25%). Di Geronimo et al. (2020) substantiated their results by discussing some remarks of their participants. Some participants noted that Dark Patterns are so widely spread and common among modern applications, that they become part of the normal interaction flow. As users are frequently being exposed to Dark Patterns, their attention for such designs is somewhat fading. In other words, users may have developed a sort of *Dark Pattern Blindness* to malicious design.

2.7.4 User's Experience with Dark Patterns

However, what happens if users *do* notice that they are being tricked? Although Dark Patterns have proven to be highly effective assets in reaching business-oriented goals, they may lead to a reduction of customer happiness in return. In addition, as users find out about dark practices, their trust in a company might diminish (Brownlee, 2016). Although various assumptions of the effect of Dark Patterns on user experience exist, remarkably, none of these assumptions are actually researched in an academical setting. More specifically, research investigating the effects of Dark Patterns on user perception or emotional appraisal are non-existent.

2.8 Why do Dark Patterns work?

The effectiveness of Dark Patterns can for most part be explained by adopting a psychological perspective: Dark Patterns often work well because they take advantage of the psychological constitution of human beings (Bösch et al., 2016). Additionally, technological factors play a role in the functioning of these malicious design strategies, as Dark Patterns often appear on digital platforms (Nevala, 2020).

2.8.1 Technological Factors

Information technology is constantly evolving, and so are the persuasive and manipulative strategies that users face. The velocity of this change often leaves the user unaware of potential threats. This phenomenon is known as *information asymmetry*. The service has more information on the user than the user has of service, causing a power imbalance (Acquisti et al., 2017).

2.8.2 Prompting System 1 Thinking

According to the Dual Process Theory of Kahneman (2011), humans have two modes of thinking: 'System 1 thinking' and 'System 2 thinking'. Whereas the System 1 thinking process takes place automatically, unconsciously, and with little effort, the System 2 thinking process takes place in a controlled, conscious and effortful way. As such, when people are presented with a stimulus, System 1 automatically and intuitively generates a response to it (Dennis et al., 2020). According to Bösch et al. (2016), Dark Patterns make use of this quick response. They exploit users' System 1 thinking to get them to make a decision which is desired by the designer and/or online service.

One example of a situation in which Dark Patterns make use of System 1 thinking has to do with the colours used in a User Interface. Generally, users associate certain colours with certain functionalities, such as green for continuing or allowing something, grey as inactive, and red as cancelling. Users use these associations to intuitively and quickly click through -for example- popups or forms. However, they might mistakenly click the wrong option if its colour differs from the users' automated thought process.

Another example relates to users agreeing to terms and conditions when creating a new account on a website. Terms and conditions are often not read, and agreement is typically made automatically and quickly (i.e., System 1 operates). This gives designers the opportunity to fill the general terms and conditions with dark ingredients, without the user knowing (Bösch et al., 2016).

A last example in which users' System 1 thinking is exploited to make users perform an action which is desired by the designer is the advertisement with the strand of hair photo shopped in it (figure 1, discussed in the Introduction). The intuitive and automatic 'System 1' reaction of users when encountering this hair will most likely be to 'swipe it off their screen'. However, since the hair is digital instead of physical, users end up swiping up on the ad, and travelling to the advertisements website.

In terms of Norman's action cycle: The user is involved in the *visceral level* of processing. The user subconsciously reacts to the advertisements aesthetics. Feelings of disgust will lead the user to perform the action of 'wiping the hair of their screen', while the unexpected result of performing this action will lead to feelings of frustration.

2.8.3 Exploiting Cognitive Biases

In addition to exploiting users' System 1 thinking to get them to perform a certain action, designers of Dark Patterns also make use of more specific psychological mechanisms. Many Dark Patterns operate by exploiting heuristics and cognitive biases in users (Mathur et al., 2019).

Heuristics are mental shortcut strategies. The application of heuristics can result in a logical fallacy based on cognitive factors: a cognitive bias. Cognitive biases can be used in either one of two ways in designing user interfaces. On the one hand, they can be countered to help people make better decisions. On the other hand, they can be exploited to influence the behaviour of people in whatever way is desired by the designer or (online) service (Gilovich et al., 2002).

Quite a few cognitive biases are being used in the design of malicious interfaces. In what follows, the most important cognitive biases used in Dark Pattern strategies will be discussed. Table 4 summarises the cognitive biases, and presents an overview of which biases are used by which Dark Pattern strategies.

Default effect The default effect is the tendency of individuals to stick with options that are assigned to them by default (Johnson et al., 2002). These individuals assume that the default setting is chosen on the basis of what is best for them. As such, they often do not even look at, let alone change, the default setting (Forbrukerrådet, 2018).

Dark Pattern strategies that make use of the default effect in their UI are *Preselection* and *Sneak into basket*. The preselection Dark Pattern exploits the default effect by, as the name already suggests, preselecting a certain option as a default. Many platforms, for example, select the option of giving consent to sharing personal data for marketing purposes as a default. People are susceptible to this preselected option, as they believe the option is aligned with their preferences.

The sneak into basket Dark Pattern exploits the default effect in users by hoping that users will stick with the products it adds to their shopping cart, often promoting the added products as 'bonuses' and 'necessary'.

Framing effect In the framing effect, individuals reach different decisions based on the same information, depending on how it is presented (Tversky and Kahneman, 1981). Equivalent information can be more or less attractive depending on what features are highlighted. A positive presentation of something emphasises the gain, whereas a negative presentation stresses the loss, heightening the effect of loss aversion (Acquisti et al., 2017).

Dark Pattern strategies that use the framing effect in their UI are *Toying with Emotions* (Brignull's Confirmshaming) and *Trick Questions*. An example of the framing effect in the Toying with Emotions Dark Pattern is when users are presented with a popup that asks for the permission to use cookies. The options users might be faced with could be "Yes, I want to ensure the best user experience" and "No, I hate websites that function well". By framing the negative option as such, users might be afraid they will lose out on something, and as a result will be more likely to agree to the cookies.

An example of the framing effect in the Trick Questions Dark Patterns is when a certain choice is framed as of it aligns with the preferences of the user, while in fact it is not.

Scarcity bias Scarcity bias is the tendency of individuals to place a higher value on things that are scarce (Mittone and Savadori, 2009). Scarcity is associated in the human brain with something positive, luxurious and exclusive. Individuals automatically assume that when something is scarce, it must be because everyone want or has already bought this product, and therefore it must be a good product. In other words, scarcity arouses peoples interests (Convertize, n.d.).

As the word 'arousal' is involved, it might be no surprise that the *Toying with Emotions* Dark Pattern uses the scarcity bias. One instance of this Dark Pattern is the countdown timer, which is a dynamic indicator of a deadline, counting down until the deadline expires. Countdown timers are often used in the check-out process of e-commerce websites. They arouse the user by letting them believe that when the timer counts all the way down to 0, a certain offer expires (although the timer actually will reset after timeout, with the offer still being valid). By giving users the feeling that the 'time to get this great offer is scarce', people will be more inclined to buy a certain product or service quickly.

Table 4: Cognitive Biases, their short descriptions, and Dark Pattern Strategies that make use of them.

Cognitive Bias	**Description**	**Used by Dark Pattern**
Default Effect	Tendency of an individual to stick with options that are assigned to them by default.	Sneak into Basket, Preselection
Framing Effect	Phenomenon that individuals may reach different decisions from the same information depending on how it is presented.	Toying with Emotions, Trick Questions
Scarcity Bias	Tendency of an individual to place a higher value on things that are scarce.	Toying with Emotions
Sunk Cost Fallacy	Tendency of individuals to continue an action if they have invested resources (e.g., time and money) into it, even if that action might make them worse off.	Hidden Costs

Sunk Cost Fallacy The sunk cost fallacy is the tendency of individuals to continue an action if they have invested resources (e.g., time and money) into it, even if that action might make them worse off.

A Dark Pattern strategy that uses the sunk cost fallacy is the *Hidden Cost* Dark Pattern. As explained in previous sections, this pattern reveals new, additional, and often unusually high charges to users just before they are about to complete a purchase, and after the user has already filled out shipping/billing information. At this point, users already feel so invested in the process, that they likely justify the additional charges by completing the purchase to not waste their effort.

2.9 Towards Ethical UI Design

Since the introduction of the term 'Dark Patterns', the ethical aspects of UI design have been increasingly gaining attention in the HCI community. A growing number of publishing addresses the unethical nature of designing for deception, with the goal of working towards more 'Ethical UI'. Although there is no widely established definition of this term, experts in the field have provided their take on it. For example, Karr (2014) stated: "I like to think of ethical things as thoughts, words, behaviours, designs, systems, and customs that are cumulatively more beneficial than they are harmful."

Many of the publications about Dark Patterns deliberately explored the techniques used by the "bad guys", for example by collecting malicious strategies, and proving their effectiveness. This approach may initially seem suspicious, as it could provide guidance for malign stakeholders. However, researchers within the field generally believe that this approach is helpful and necessary, for a number of reasons.

First of all, a detailed analysis and documentation of Dark Patterns allows for a better understanding of the underlying concepts and mechanisms threatening the users. Second, research on Dark Patterns fosters awareness among users, and makes it easier for users to identify such malicious patterns in the wild. And last, research on Dark Patterns can be used as a starting point for the development of countermeasures, and even (governmental) regulations (Bösch et al., 2016).

The "bad guys" approach seems to work, as research on Dark Patterns already has led to some change -or at least awareness- in the world of UI design. For example, in January 2021 the Norwegian Consumer Council filed a legal complained against the American e-commerce company Amazon. The Council analysed the cancellation process for Amazon Prime, the streaming service of Amazon, and found that consumers who want to leave the service are faced with a large number of hurdles. As such, users get confused and are tricked into continuing paying services they do not need or want. In other words, Amazon uses the *Roach Motel* Dark Pattern to manipulate users to stay subscribed. Based on the complained of the Norwegian Consumer Council, 16 other consumer organisations in Europe and the United States are now taking action against Amazon. Additionally, they will each ask their respective consumer authorities to investigate the use of Dark Patterns in their countries (Forbruker Radet, 2021).

Awareness of Dark Patterns also seems to have created a shift in the mindset of companies that could potentially benefit from Dark Patterns themselves. In recent years, an increasing number of blog posts appeared on companies websites, discussing why *not* to use Dark Patterns. Some of these blogs are even written on e-commerce platforms, where individuals can create their own online shops. For example, the e-commerce platform Shopify features a blog on its website which discusses 12 Dark Patterns and why you -as an entrepreneur using their platform- should never use them. They mention, amongst others, that Dark Patterns erode trust and take advantage of potential consumers (Shopify, 2019).

2.10 Aim of this study

The current study aims to contribute to this positive development within the field of interface design. To decide upon the scope of the current study, both the preceding literature research and the opinion of Harry Brignull -an expert in the field of Dark Patterns- were taken into account. This has led to the identification of three clear gaps in the existing body of knowledge.

As such, the current research is threefold:

2.10.1 Measuring the effect of Dark Patterns on User Experience

First of all, this study aims to contribute to the existing research upon the topic of Dark Patterns by gathering new insights -by looking at Dark Patterns from a new point of view. While previous research has mostly focused on the categorisation, prevalence and effectiveness of Dark Patterns, the current study takes a more user-centric approach.

As already mentioned in section 2.7.4, little research has been performed on how users perceive or experience Dark Patterns (e.g. how they *feel* about them whenever they notice them). Based on the fact that Dark Patterns trick the user into doing things they did not intent to do, one might think that a user's experience with a product or service is negatively influenced as the user encounters -and recognises- a Dark Pattern in a user interface. In other words, a user might experience negative emotions, attitudes or feelings as a result of encountering a Dark Pattern.

To investigate whether this assumption is true, the current study will perform a (quantitative) experiment. The results of the experiment will provide an answer to the following research question:

*RQ*1 *"Does the use of Dark Pattern strategies in UI design influence a user's experience with a service or product in comparison to these malicious strategies being absent?"*

2.10.2 Assigning a Severity Score to Dark Patterns

The subjective ethical nature of Dark Patterns can be addressed if we can determine how significantly a pattern affects the users and their experience of the product or service. Although all types of Dark Patterns are deployed with the same (malicious) intent in mind, the different types of Dark Patterns may vary on how strongly they affect users.

As such, the second aim of this study is to measure the severity -or 'darkness'- of the various types of Dark Patterns. Ideally, each of the Dark Patterns mentioned in Table 3 will be assigned a certain severity score. In what follows, this score will be referred to as the 'Dark Pattern Darkness Score' (DPDS).

2.10.3 Creating a 'System Darkness Scale'

As (malicious) designers may use different amounts and combinations of the various Dark Pattern types, different services or products will have different overall severity scores.

As such, the last aim of this study is to create a measure of the darkness of a service or product as a whole. The current study takes the first steps towards creating a 'System Darkness Scale' (SDS). The goal of the SDS is to quickly and easily determine the 'darkness' of a system, service, or product.

In a certain way, the SDS will be comparable to the System Usability Scale (as discussed in section 2.2.2). However, whereas the SUS provides an insight into how *usable* a certain product or service is, the SDS will provide an insight into how *dark* a certain product or service is.

3 Research Methodology

This section elaborates upon the set-up of -and all the methods used within- the current study. As mentioned in subsection 2.10, this research aims to contribute to the existing body of knowledge in three ways. In order to provide a clear insight into each of the three objectives, the objectives were translated into three different studies. The current section will first discuss the general and overarching aspects of the three studies (section 3.1). Thereafter, we will dive into the specifics of each study. An explanation upon the performance of an experiment on a user's experience with Dark Patterns in UI design is provided in section 3.2. Section 3.3 entails the details of how the first steps towards creating a 'System Darkness Scale' were taken. The Methodology section ends with discussing the approach to assigning severity scores to the various types of Dark Patterns (section 3.4).

3.1 General Aspects of the Study

3.1.1 One Study, Three Study Parts

Although each of the research objectives (as described in subsection 2.10) was translated into its own study, the researcher created one overarching study in which participants were guided along each of the three studies during one session. In other one big study was created, consisting of three 'study parts'.

This approach has several advantages. First of all, in order to provide reasonable validity to the outcome of the studies, a large sample size is required. By performing one big study instead of three separate ones, the process of recruiting a large sample of participants became much easier. Participants only had to be recruited once -for actually partaking in all three studies. Each participant, therefore, added to the sample size of each of the three studies.

If the researcher had decided to run each of the studies separately, chances are that the same participant would not have participated in all three of them. In this scenario, participants would most likely have encountered *research fatigue* -a process or state in which individuals tire of engaging in research or resist and avoid participation in any further research (Clark, 2008). This would have resulted in the need to recruit more participants to reach the (same) desired sample size for each of the three studies.

This brings us to the second point. In the case of the current research, the order of performing the three study parts is of great importance, as one study can prime participants knowledge for the next one (this will be explained in section 3.1.3). Whereas one can control the order of the three study parts by creating one overarching study, this is not possible -or at least more difficult- when running the studies separately. One could solve this issue by not recruiting the same participant for all three studies, however, then we will face the issue of having to recruit way more participants again.

The last advantage of running one overarching study is related to time. There are some elements of every research that cannot be omitted. Think about a 'general questionnaire', which gathers some of the participants descriptives, and a 'consent form', which informs the participants about the objectives of the research and their personal rights. By creating an overarching study, the participants only had to encounter -and fill out- these elements once (instead of three times).

3.1.2 Online Experimental Environment

In order to control the order of the study parts, and guide participants through the whole study seamlessly, an online environment was created. The online environment took the form of a website and consisted of several web pages. By performing a specified action on a certain web page (for example clicking a 'next' button at the bottom of the page), the participant was guided to the next web page in line.

3.1.3 Outline of Study

The final overarching study consisted of several elements. An overview of the order of all these elements is presented in figure 18. Screenshots of the web pages representing each of the elements can be found in Appendix A.4. In what follows, each of the elements will shortly be discussed.

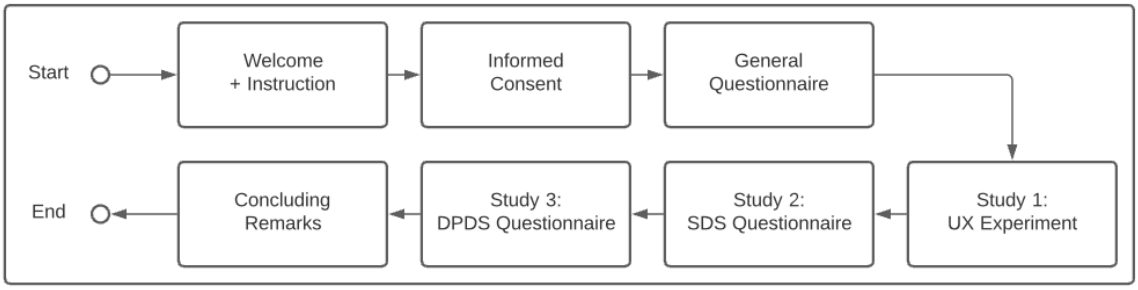


Figure 18: Outline of Study.

Welcome and Instruction The first thing that appeared to the participant when entering the online environment was a welcome and introduction text. The text thanked participants for partaking in the study, and provided participants with a short general overview of what they could expect from it. Participants were asked to (1) perform the study on a laptop (not on a mobile device), (2) be seated in a quiet environment, and (3) refrain from spreading information about the study. Furthermore, participants were told that the study should be fully self-explanatory, but that they should not hesitate to contact the researcher about any questions they might have. Participants were also asked to choose a username (and stick to this username during the whole study), in order for the researchers to be able to distinguish participants from each other while still preserving anonymity.

Informed Consent The second web page included a consent form. The consent form informed the participants about the data gathering of the study. It stated that all the data would be anonymously processed in the final outcome of the study. Additionally, the consent form included some statements about the personal rights of the participant, such as the right to withdraw from partaking in the study at any time, for any reason.

General Questionnaire The third web page included the general questionnaire. Within this questionnaire, participants were asked to fill out some general information, such as their date of birth, gender, and highest level of education. The general information needed to be filled out in order for the researchers to be able to describe the final sample of participants.

Study 1: UX Experiment After the general part of the study was completed, participants were forwarded to a web page explaining the tasks for the first study. Participants had to perform a shopping task on a (fake) shopping website. While performing this task, half of the participants encountered Dark Patterns. At several points during the shopping task, participants were prompted with a pop-up, which asked them to rate their experience with the design of the shopping website. At the end of the shopping task, participants were also asked to indicate whether they would recommend the website to friends. A more detailed explanation upon Study 1 is provided in section 3.2.

Study 2: SDS Questionnaire After participants finished Study 1, they were guided to the explanation page of Study 2. A link to Google Forms brought participants to a questionnaire containing 25 statements related to the Darkness (or Brightness) of the system they had just interacted with (e.g., the shopping website). Participants had to indicate whether they agreed or disagreed with each of the statements on a 5-points Likert Scale. A more detailed explanation upon Study 2 is provided in section 3.3.

Study 3: DPDS Questionnaire Within the last study part, participants were presented with 17 situations in which Dark Patterns were involved. The situations were described in text and supplemented with an example image. Participants were asked to imagine themselves being in the situation at hand, and indicate how they would feel about the situation. Participants also had to indicate how frequently they had experienced a similar situation. A more detailed explanation upon Study 3 is provided in section 3.4.

Concluding Remarks After finishing the last study part, participants were guided to the final web page. The final web page of the study thanked participants for partaking in the study, and explained shortly what the whole study was about. It also provided participants with the opportunity to fill out their name and e-mail address to receive the results of the study.

As was mentioned in the previous subsection already, the order of the three study parts was of great importance, as one study part could prime the knowledge of participants for another study part. On the other hand, sometimes the content of one of the study parts came in helpful for another study part. For example, the 'Bright' and 'Dark' shopping websites used in Study 1 (UX experiment), could be used as 'evaluation material' for Study 2 (SDS questionnaire). As such, Study 2 needed to be performed after Study 1. And as Study 3 (DPDS) revealed all of the Dark Pattern types, and we did not want to prime participants with those in Study 1 and 2, this Study should be the last one in line.

3.1.4 Devices, Software & Tools

Due to the COVID-19 restrictions, participants performed the study on their own laptops. As such, diagonal sizes and resolutions of laptop displays may have differed amongst participants.

The online experimental environment was built on the personal domain of the researcher. HTML5, CSS3, PHP7 and JavaScript were used to create the desired look and functionality of the environment. The data of the general questionnaire, data gathered in Study 1, and names & e-mail addresses of participants who wanted to receive the results of the study were written to a txt file on the server. The data of Study 2 and 3 were gathered using Google Forms, and automatically transferred to an Excel file.

All HMTL, CSS, PHP and JS files used in the current study can be found on GitHub:
https://github.com/xxkristi/DarkPatterns.git

3.1.5 Participants

Since part of the design of the current study is a between-subject design (Study 1 & 2), a relatively high number of participants was needed to be able to validly compare the two experimental conditions. Participants were selected based on accessibility and, due to the COVID-19 restrictions, recruited over the internet. There were no age-, gender- or familiarity criteria, as a wide variety of users of services containing Dark Patterns exist. As such, this research did not discriminate between them.

Section 3.1.3 already shortly described each of the three study parts. In what follows, we will describe the study parts in more detail, starting with Study 1: Performing an Experiment on User Experience.

3.2 Study 1: Performing an Experiment on User Experience

In order to answer RQ1: "Does the use of Dark Pattern strategies in UI design influence a user's experience with a service or product in comparison to these malicious strategies being absent?", a between-subjects experiment was conducted.

3.2.1 Experimental Tasks

Within the experiment of Study 1, participants needed to perform three tasks: A Shopping Task, a UX Rating Task, and a Referral Rating Task. All three tasks needed to be performed on an e-commerce website, which was created by the researcher. The e-commerce website resembled an online "gesture" shop, selling various types of postcards, bunches of balloons and flower bouquets.

The decision to use an e-commerce website as "the system" for this experiment was made for several reasons: The main reason being that the experiment took place in times of COVID-19. As such, participants needed to be able to participate in the experiment on their own. E-commerce web shops seemed to be the perfect platform for performing the experiment online and at distance, as they are easily accessible and do not require additional explanation.

In what follows, each of the experimental tasks will be explained.

Shopping Task The shopping task was formulated to the participant as follows: "You want to send four of your friends a postcard to let them know you still think of them in times of COVID-19 and you want to meet them online soon. Find the best deal on COVID postcards in the shop, add them to your shopping basket, pay for them, and make sure they're coming your way! You don't need stamps or envelopes, as there's still lots of them in the drawer of your closet."

Participants were told that they could freely browse the shopping website, as long as they fulfilled the shopping task. They were also asked to behave as if they were really involved in the act of online shopping, and match their behaviour accordingly.

The e-commerce website helped the participants complete the shopping task, by guiding them through the steps generally involved during shopping: Searching for a product, deciding whether a product meets the requirements (in this case, the requirements of the task), adding the product of choice to the shopping basket, checking the details of the order, and paying the product.

UX Rating Task At several points during the shopping task, participants were prompted with a full screen pop-up asking them to rate their experience with browsing the page that they had just visited. Participants performed the rating on a 5-point Likert scale, ranging from 'Very Unsatisfied' to 'Very Satisfied' (figure 19).

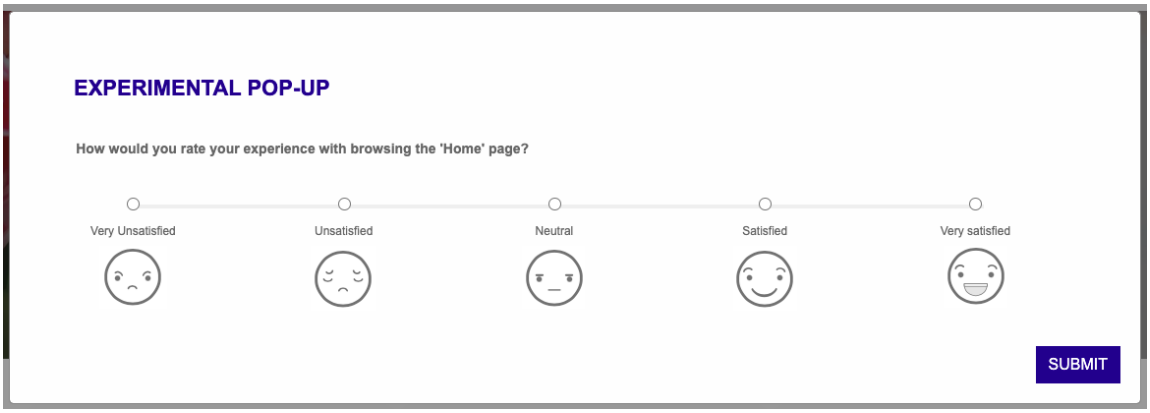


Figure 19: UX Rating Pop-up

Referral Rating Task As soon as the participant reached the last page of the shopping website, and placed his order, a final pop-up appeared asking the participant how likely they were to recommend the e-commerce website to a friend. Participants needed to indicate this using a 10-point Likert scale, ranging from 'Not at all likely' to 'Extremely likely'. Participants were also provided with the opportunity to explain their score (figure 20). The latter was not required, however.

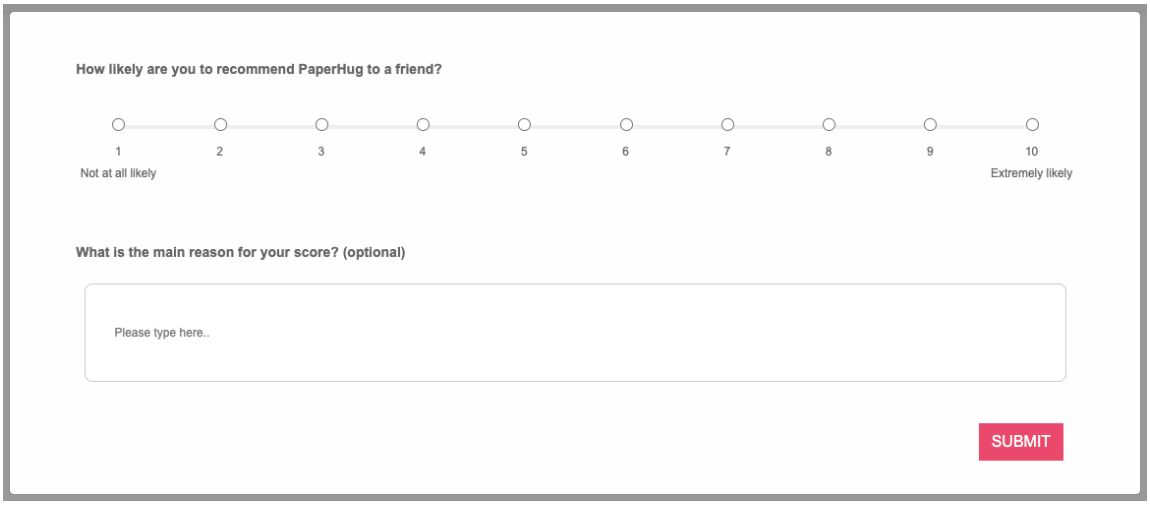


Figure 20: Referral Rating Pop-up

The shopping task was necessary to provide some context to the experiment. It made sure every participant had the same goal in mind. By specifying a task, all participants were encouraged to visit the same pages of the website. This way, the researcher could better control the content participants encountered within the interface of the web shop.

Although the shopping task only provided some context for the participants, the UX rating task was where the most important data was gathered. This data is the data that needed to be analysed in order to answer the research question of Study 1.

The referral rating task was conducted in order to gather some more insights into the participant's thought processes. Whereas the goal of the UX rating task was to gather some insights concerning participant's experience with the individual web pages of the e-commerce web shop, the goal of the referral rating task was to capture the 'overall experience' participants had with the web shop. By asking participants to write down some textual explanation of their score, the researchers also hoped to gain some insights into the likes and dislikes of the (interface design of) web shop as well.

3.2.2 Experimental Design

As already mentioned in section 3.2.1, participants needed to interact with an e-commerce website. As we want to measure whether Dark Pattern strategies have an effect on User Experience, two versions of an e-commerce website were created. Whereas Version A (control condition) *did not include* Dark Patterns at all, in Version B Dark Patterns *were included* in most of the pages that participants needed to visit to complete the shopping task.

Mathur et al. (2019) identified the following Dark Pattern strategies to be often used in e-commerce websites: 'Price Comparison Prevention', 'Hidden Costs', 'Sneak into Basket', 'Hidden information', 'Preselection', 'Toying with Emotions', 'Trick Questions', 'False Hierarchy', and 'Disguised Ads'. As such, these malicious strategies were selected to be implemented in the Dark version of the e-commerce website.

An example of the differences in interface design between the two experimental conditions is provided in figure 21. Appendix A.4 shows the implementation of all of the Dark Pattern strategies in the web pages of the B version, and compares these pages with the design of the pages in the A version.

Participants were randomly assigned to either the A or B version of the experiment. The assignment took place at the explanation page of Study 1. This page consisted of a textual elaboration on the various tasks, and a button which participants could click after they read the elaboration. An invisible counter was tied to the button. Each click on the button resulted in the counter adding +1 to its initial value. If the value of the counter was an even number, the participant was directed to the A version of the e-commerce website. If the value of the counter was an odd number, the participant was directed to the B version of the-commerce website.

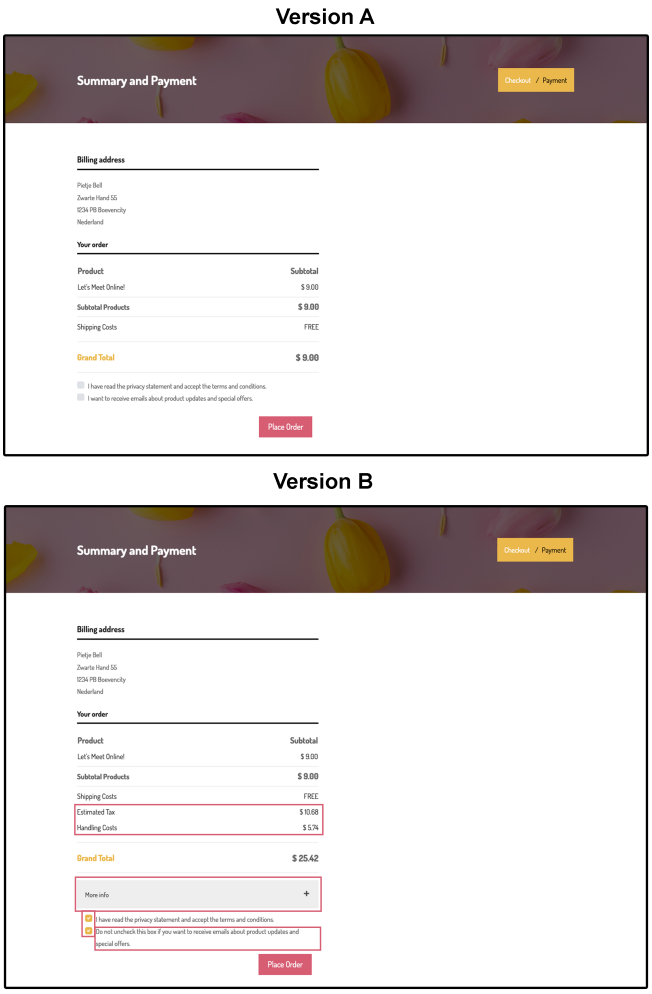


Figure 21: Example of different UI design in Version A (top) & B (bottom) of the 'Place Order' page. In Version B, the 'Hidden Costs', 'Hidden Information', 'Preselection', and 'Trick Question' Dark Patterns are included in the interface.

The e-commerce web shops of Version A and B of the experiment can be found at:

- https://www.kristibergman.nl/Ahome.html
- https://www.kristibergman.nl/Bhome.html

3.2.3 Research Material

Website Design

The two versions of the e-commerce website both consisted of the same set of pages: a 'home' page, an 'about' page, several 'product categories' pages, several 'product overview' pages, several 'product details' pages, a 'shopping cart' page, a 'checkout' page, and a 'place order' page. Therefore, navigating between the pages worked the same for both Version A and B of the website.

In what follows, a short description is provided for each of the pages:

- The 'home' page presented participants with a banner, showing the focus of the e-commerce web shop -specifically 'selling thoughtful surprises'. Beneath the banner, three shortcuts to 'product category' pages were provided -all featuring a different surprise (e.g., balloons, postcards & flowers).
- The 'about' page contained a short text explaining the goals and objectives of the web shop.
- The 'product categories' pages consisted of a list of categories related to the selected surprise. For example, when selecting the 'Postcards' surprise, the 'product categories' page showed a list containing 'Birthday cards', 'COVID-19 cards', 'Thank you cards', and 'Valentines cards'.
- The 'product overview' pages showed participants the various products the shop offered within a category (e.g., different designs of cards; single cards and sets of cards).
- The 'product details' pages presented participants with more details on a selected product, such as a short description, price, and reviews. The page also featured an input field where participants could indicate the desired quantity of the product and an 'add to cart' button.
- The 'shopping cart' page presented all the products that were added to the cart during the shopping process. Participants were able to delete products from here if they wanted to. The shopping cart page additionally revealed information regarding prices (e.g., subtotals and totals).
- The 'checkout' page showed participants a form where they needed to fill out billing information. At this page, participants also had to indicate their preferred shipping- and payment method. The page also provided participants with a short summary of their order.
- The 'place order' page showed participants a summary of the billing information they filled out on the 'checkout' page. It also provided participants with an updated summary of the costs. Last, it presented participants with two check boxes asking participants to agree to the privacy policy and whether they wanted to receive promotional offers.

Screenshots of the various pages for both Version A and B of the experiment are provided in Appendix A.4.

For completing the shopping task, participants had to visit the 'home' page, the 'postcards categories' page, the 'postcards overview' page, (at least) one of the 'postcards details' pages, the 'shopping cart' page, the 'checkout' page, and the 'place order' page. In what follows, these seven pages will be referred to as *task-related pages*. Figure 22 presents a visual overview of the task-related pages.

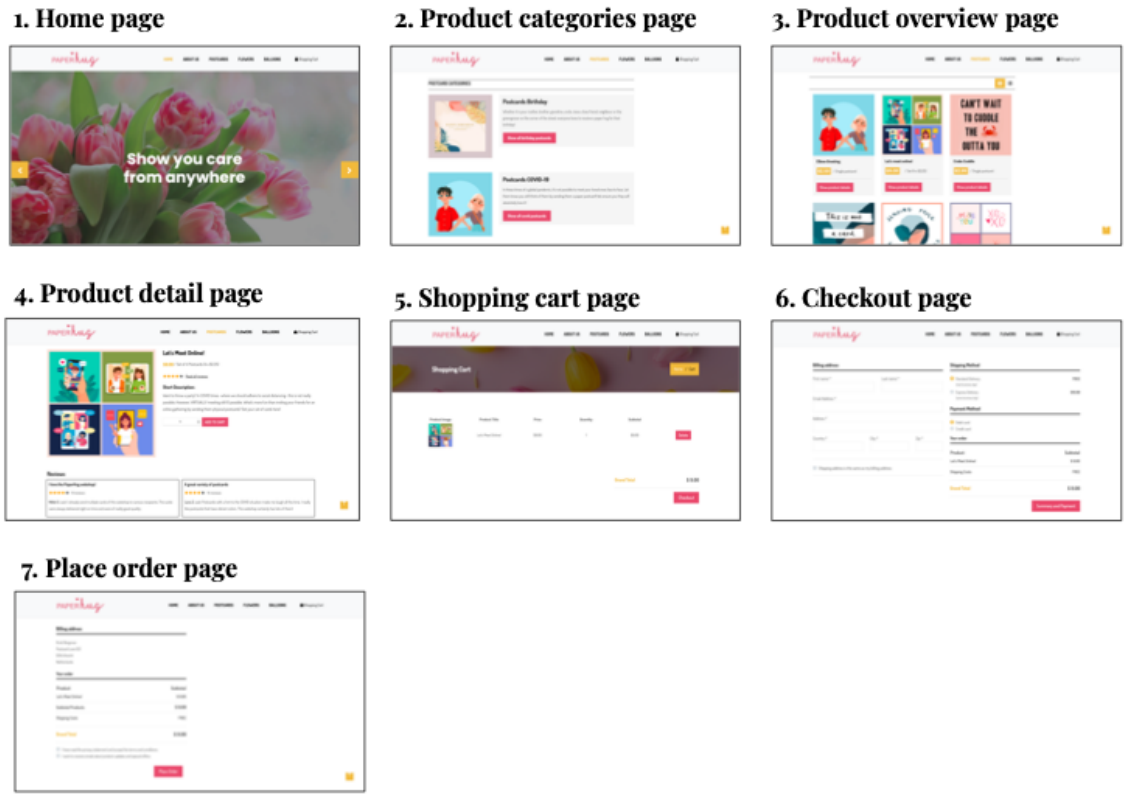


Figure 22: The seven task-related pages, which participants had to visit in order to complete the shopping task. Within the B version, Dark Patterns were present in the task-related pages.

Table 5 shows which Dark Patterns were included in each of the task-related pages.

Table 5: Implementation of Dark Patterns in the various Task-Related Web pages of Version B. The 'checkout' page is not mentioned here, as this page did not include Dark Patterns in both versions.

Page	**Dark Pattern**	**Instantiation**
Home (Cookie consent)	Hidden Information	Info about the various cookies is hidden in second screen of pop-up (manage cookies).
	Preselection	All cookies are automatically accepted on the second screen of the cookie consent.
	False Hierarchy	One button has visual precedence over another button (color, size, style).
Product Categories (Offer pop-up)	Toying with Emotions	Emotional wording in pop-up offering 20% off next order.
	False Hierarchy	One button has visual precedence over another button (color, size, style).
Product Overview	Price Comparison Prevention	Direct comparison of prices for single / sets of cards is not possible.
	Toying with Emotions	Show scarcity / high demand messages for all of the products.
Product Details	Disguised Ads	Ads look similar to customer product reviews.
Shopping Cart	Sneak into Basket	Stamps are automatically added to the shopping cart, without the user knowing.
	Toying with Emotions	Countdown timer for limited reservation of shopping order.
Place Order	Hidden Costs	Additional fees & taxes are added at last stage of checkout process.
	Hidden Information	Checkbox concerning receiving promotional emails is hidden in 'more options' tab.
	Preselection	Checkbox concerning receiving promotional emails is selected as a default.
	Trick Questions	Double negatives in the text next to the checkbox about receiving promotional emails.

Although the other pages were not necessarily important for completing the shopping task, participants *were* able to browse these pages. The pages of the 'surprises' other than postcards (e.g., balloons and flowers) also did have full functionality. As such, a participant could, for example, add a bunch of balloons to his shopping cart.

The pages unrelated to the shopping task also contained Dark Pattern strategies (in the B version) in order for the web shop to look consistent, and not differentiate between using or not using Dark Patterns for the different types of products. For example, the high-demand messages ("Bought 11 times in the last 24h!") were present on the 'product overview' page of the postcards, as well as on the 'product overview' page of the flowers and balloons.

While designing the website, the researcher closely studied the look and feel of other e-commerce websites selling postcards, such as Hallmark and Greetz. This was important in order to give participants the feeling as if they were shopping for postcards on an *actual* website.

Web Shop Pop-up Design

During completion of the shopping task, participants were presented with three pop-ups from the web shop itself. The researcher built a 'cookie modal', an 'offer modal' and a 'referral modal'.

- The cookie modal appeared to the participants right after they opened the home page of the shop. The modal looked different for the A and B version (as Dark Patterns were used in the modal of the B version), but basically asked participants to set their preferences concerning cookies.
- The offer modal was presented to the participants when they entered the 'postcards categories' page. The modal offered participants a 20% discount on their next order. In order to receive the discount, participants had to fill out their e-mail address.
- The referral modal was presented at the end of the shopping task, right after participants placed their order. As already explained before, the referral modal asked participants whether they would recommend the shopping website to their friends (and why / why not). As already mentioned before, filling out the referral modal was one of the experimental tasks.

All three modals were designed to match the colours and style of the shopping website.

Although the functionality of the cookie- and offer modal seemed to work from the participants' perspective, the preferences the participants indicated were not stored. Implementing such functionality was beyond the scope of this research.

The input of the referral modal, however, *was* stored. As such, the referral modal could have been classified as an experimental pop-up as well (see next paragraph). The researcher decided not to do this, as referral pop-ups are general practice in actual shopping websites; they are often used to gather the opinions of (potential) customers.

Experimental Pop-up Design

At certain points during their interaction with the shopping website, participants were presented with an experimental pop-up (e.g., the UX pop-up, figure 19). The pop-up appeared each time a participant clicked an element on a task-related web page (e.g. tried to navigate to another page). For example, when a participant clicked on one of the postcards items on the task-related '*postcards* overview' page in order to reach the 'details' page, the pop-up was presented.

However, when a participant was browsing the '*flowers* overview' page and clicked on one of the flower items in order to reach its 'details' page, the pop-up was not presented.

The UX pop-ups disappeared directly after the participant filled out how he or she felt towards the user interface at hand. Both these design decisions were made in order to keep impairment of the shopping task flow to a minimum.

Another design decision that was made is related to the look and feel of the pop-ups. The UX pop-ups were designed in such a way that they mismatched the colours and style of the shopping website. As such, participants could distinguish between 'web shop content' and 'experimental content'. This way, the researcher prevented participants from taking into account the many UX pop-ups in their Referral Rating. In other words, the researcher prevented participants from scoring their overall experience with the web shop a '4', because "There were too many pop-ups".

The last design decision has to do with the way the 5-point 'Satisfaction' Likert scale in the UX pop-up was presented to the participants. Next to the textual labels ('Very Unsatisfied', 'Unsatisfied', 'Neutral', 'Satisfied', and 'Very Satisfied'), smiley faces representing the textual 'states' were presented to the participant. The goal of using these smileys was to provide participants with a visual representation of the various answering options, making it easier and quicker for participants to fill out the Likert scale question. The smiley faces used in the UX pop-up of the experiment were retrieved from the 'What do you think about our Webshop?' pop-up presented to users of the "gesture" web shop Greetz.

Pictural Content

The pictures of the products presented on the e-commerce website were obtained from the website Freepik, which offers free-to-use pictures. Pictures were selected based on their resemblance with each other, as participants needed to get the feeling the e-commerce web shop had photographed each of the products itself. Additionally, pictures were selected based on the theme of the category they were to be placed in (e.g., for Postcards: 'birthday', 'COVID', 'thank you', 'valentine').

3.2.4 Procedure

Before the start of the experiment, participants were presented with a textual elaboration on all three tasks. After they had read the elaboration, participants could click a button with the text "Start Experiment". When clicking the button, the experiment started, and participants were automatically redirected to the home page of either the A or B version of the e-commerce website.

The experiment ended after participants placed their order and filled out the Referral Rating pop-up. As soon as this happened, they were redirected to the explanation page of Study 2. This page told them that they had finished the first Study, and provided them with a textual elaboration on the next one.

3.2.5 Variables

Within Study 1, we want to provide an answer to the question "Does the use of Dark Pattern strategies in UI design influence a user's experience with a service or product?". A such, the independent variable within the experiment was the 'presence of Dark Patterns'. The dependent variable measured by the experiment was 'User Experience'. This variable could not be measured in and of itself, however. Therefore, two measurable variables were recorded for each of the participants:

- The 'UX Rating Pop-up' measured User Experience by the participants' Level of Satisfaction. After interacting with each of the task-related pages, participants were asked to rate their experience with the UI design of the particular page. Participants performed the rating on a 5-point Likert scale, with (1) representing 'Very Unsatisfied', and (5) representing 'Extremely Satisfied'. Thus, the lower the Level of Satisfaction, the more negative the User Experience. Or, the higher the Level of Satisfaction, the more positive the User Experience.

- The 'Referral Rating Pop-up' measured User Experience by the participants' Recommendation Score. At the end of the shopping task, participants were asked to indicate whether they would recommend the web shop to friends. Participants performed the rating on a 10-point Likert scale, with (1) representing 'Not at all likely' and (10) representing 'Extremely likely'. Thus, the lower the Recommendation Score, the more negative the User Experience. Or, the higher the Recommendation Score, the more positive the User Experience.

Whereas the UX rating pop-up measured the User Experience for each of the pages relevant to the shopping task separately (e.g., 'home', 'postcard categories', 'postcard overview', 'postcard details', etc.), the Referral rating pop-up measured the User Experience for the web shop as a whole.

3.2.6 Hypotheses

As already mentioned before, participants partaking in the B version of the experiment encountered Dark Patterns in *most* of the steps leading towards completion of the shopping task, whereas participants partaking in the A version of the experiment did not encounter Dark Patterns at all.

In order to answer the research question, hypotheses were formulated for each measured variable.

Level of Satisfaction
The researcher foresees that the use of Dark Pattern strategies in the UI design of a web page (Condition B) has a negative influence on a user's experience relative to the UI design of a web page in which malicious design strategies are absent (Condition A).

In terms of the measurable variable, the hypothesis can be formulated as follows:

*H*1 The use of Dark Pattern strategies in the UI design of a web page (Condition B) leads to a decrease in the Level of Satisfaction relative to the UI design of a web page in which malicious design strategies are absent (Condition A).

Recommendation Score
The researcher expects that the use of Dark Patterns in the UI design of a website as a whole (Condition B) has a negative influence on a user's experience relative to the UI design of a website in which no Dark Patterns are used (Condition A).

In terms of the measurable variable, the hypothesis can be formulated as follows:

*H*2 The use of Dark Pattern strategies in the UI design of a website (Condition B) leads to a decrease in the Recommendation Score relative to the UI design of a website in which malicious design strategies are absent (Condition A).

3.2.7 Data Preparation

UX Pop-ups
As this experiment, on one hand, aims to measure User Experience at various points in time during an interaction task, Customer Journey Maps (as discussed in section 2.2.4) seem to be a valuable tool in analysing the obtained data. Whereas the various steps in the shopping task can be regarded as the various Journey Stages in the Customer Journey Map (horizontal axis), the Level of Satisfaction at every stage can be mapped as the User Experience journey of the participant (vertical axis).

By applying the Customer Journey Map as a tool to the journey of each participant, for each participant, we can create a visualisation like the one in figure 23.

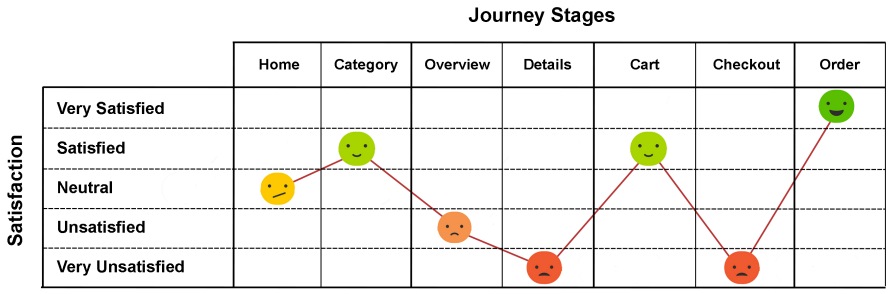


Figure 23: Customer Journey Map used for analysis in the Experiment.

By mapping the customer journeys of participants who completed different conditions, we can see whether there are differences in the user experience ratings of participants for the steps that did (Condition B) or did not (Condition A) include Dark Pattern strategies (figure 24).

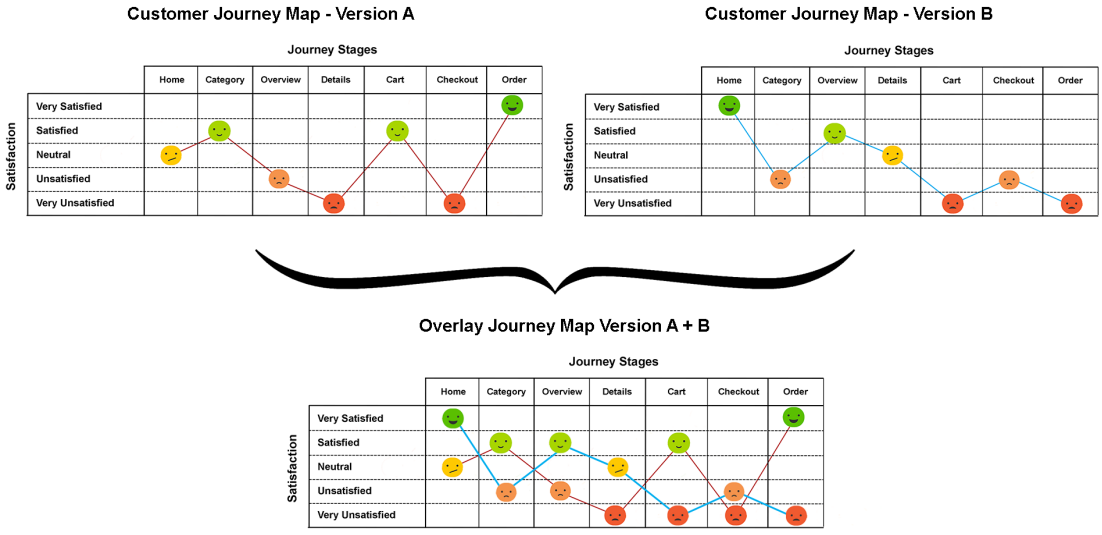


Figure 24: Mapping the Customer Journeys of Version A and B.

The customer journey maps, however, will most likely only be useful as a visual representation of the final results. To see whether a real difference exists between the various pages of Versions A and B of the e-commerce websites, statistical tests will be used.

In order to perform some statistical tests with the data, and create visualisations like the ones above, the data of the UX pop-up needed to be prepared and cleaned.

For each participant, the input of each of the seven UX pop-ups was formatted as [Version-Page-Rating], and was automatically saved to the server as a .txt file. As such, for each participant, the .txt file contained data which looked like this: "A-home-satisfied, A-categories-neutral, etc.".

All .txt files were loaded into the statistical tool R, and put together into one data frame. This data frame was formatted in a way that each row represented a participant, and each column represented a web page. The first column of the data frame represented the Version the participant was in.

The textual UX 'Satisfaction Ratings' in the cells of the data frame were converted to numerical values in order to be able to perform statistical tests with the data (figure 25).

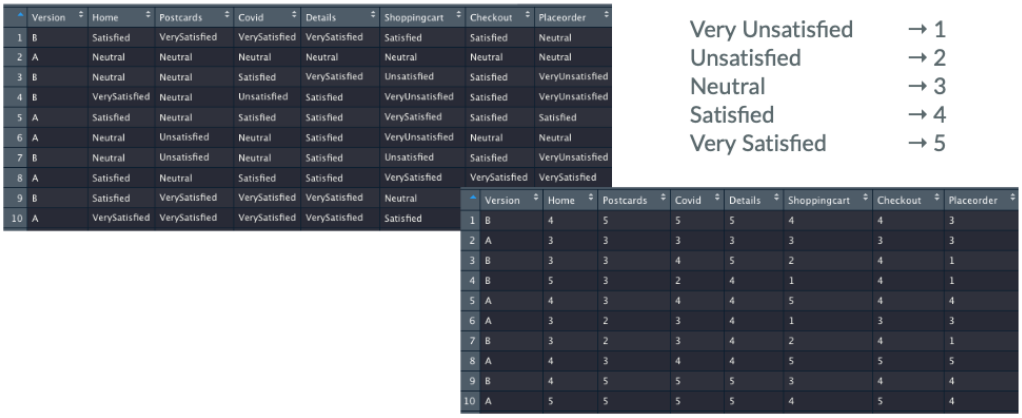


Figure 25: Converting satisfaction ratings to numerical values.

Referral Pop-ups

The data of the referral pop-ups was saved as a .txt file on the server as well. The input of the referral pop-up was formatted as [Version-Rating-Explanation]. As such, for each participant, the .txt file looked like this: "B-3-The costs were two times as high as the postcards itself".

The .txt files of the Referral pop-up were loaded into R as well, and put together into another data frame. Within this data frame, each row represented a participant. The columns in the data frame represented 'Version', 'Rating', and 'Explanation'. For the referral pop-up data, none of the data needed to be converted into another type of value.

3.3 Study 2: Creating a 'System Darkness Scale'

The second objective of the current study is to take the first steps towards creating a System Darkness Scale (SDS), which can be used to evaluate the *Darkness* of a certain system.

In developing the SDS, this study will take the same approach as was used for developing the System Usability Scale, which captures the *Usability* of a system (Brooke, 1996).

The final SDS will consist of a questionnaire containing several statements to which a user of a certain system can respond to on a 5-point Likert Scale. The responses to the statements will lead to a final score which represents the *darkness* of the system.

3.3.1 Selecting Items for the SDS

In order to make sure the scale validly captures the *darkness* of a particular system, it is important that the questionnaire items are carefully selected in the first place. The technique that is often used for selecting the items for a Likert Scale questionnaire (and that was used for selecting the items of the SUS as well) consists of multiple steps:

- The first step is to pick two situations (in this case: systems) that are at the extreme ends of the spectrum of the attitude that needs to be captured. In case of the SDS, we are interested in the attitudes towards (malicious) interface design. As such, the extreme ends of the spectrum were decided to be a "really dark system" (e.g., a system with lots of severe Dark Patterns) and a "really bright system" (e.g., a system with no Dark Patterns at all).

- The second step is to create a list of potential questionnaire items. The potential items need to cover a variety of aspects of the feature that one wants to measure -in this case 'darkness' of the system. An example of a potential statement for the SDS could be *"The system tricked me into sharing information I did not intend to share."*

- As the extreme ends of the spectrum and potential questionnaire items are decided upon, a study needs to be performed. For both extreme ends of the spectrum, a sample of participants is asked to indicate whether or not they agree to each of the potential questionnaire items on a 5-point Likert scale. In case of the SDS, for both the "really dark system" and the "really bright" system, participants were asked to give ratings to the *exact same list* of potential questionnaire items.

- The last step in selecting items for the final questionnaire is to evaluate the responses on the potential questionnaire items. Given the large pool of potential questionnaire items, there will be some items that provoke extreme agreement or disagreement among the respondents. For example, the statement *"The system tricked me into sharing information I did not intend to share."* could lead to extreme agreement for the "really dark system", whereas it could lead to extreme disagreement for the "really bright system". Statements that lead to extreme opposite responses are the ones

that should be included in the final questionnaire. Items where there is ambiguity are not good discriminators of attitudes, and therefore should not be included in the final list of items.

In what follows, each of the steps for selecting the questionnaire items for the SDS will be explained in more detail. Thereafter, the methodology concerning 'how to score the final SDS?' will be described.

3.3.2 The Extreme Ends of the Spectrum

As already mentioned in the previous section, this study is interested in the attitudes of people towards (malicious) interface design. In order to measure the attitudes of people on this whole spectrum, two extreme opposite systems were selected: A "really dark system" with lots of severe Dark Patterns, and a "really bright system" with no Dark Patterns at all.

As it seemed impossible to find two (existing) systems that are exact opposites of each other, the researcher decided that the systems needed to be self-designed. Another reason which added to this decision is the fact that in existing systems, there would be less control over -for example- what the participant sees when interacting with the system. In a self-designed system, controlling these kinds of aspects would be much easier.

As for Study 1 a "dark" (condition B) and "bright" (condition A) system were developed already, the researcher decided to use these systems as 'the extreme ends of the spectrum' for Study 2. As such, the e-commerce website containing Dark Patterns became the "really dark system", and the e-commerce website containing no Dark Patterns became the "really bright system".

As participants were designated a dedicated task in Study 1, they all saw the same set of web pages. Therefore, a good comparison between the two systems could be made.

3.3.3 List of Potential Questionnaire Items

The potential list of questionnaire items needed to cover a variety of aspects of the feature that the study wants to measure. As such, the researcher wrote down a variety of aspects related to 'darkness' and 'dark patterns'. Aspects included (amongst others): 'trickery', 'deception', 'evil intentions' and 'business-centric approach'. Next, based on all the aspects that were derived, the researcher formulated 25 potential questionnaire items. The list of potential questionnaire statements is provided in table 6.

Note that about half of the statements is formulated in a "dark" manner, whereas the other half of the statements is formulated in a "bright" manner. This was done in order to prevent response biases caused by respondents not having to think about each statement. By alternating "dark" and "bright" items, the participants had to read each statement carefully, and had to make an effort to think whether they agreed or disagreed with it.

Table 6: List of Potential Questionnaire Items.

Item	Statement
01	The system tricked me into performing certain actions that I did not intend to do.
02	I was repeatedly interrupted by the system without being able to stop the interruption.
03	I could respond to the system as I desired at all times.
04	The system obstructed me in performing certain actions.
05	I could perform every action that I wanted to perform.
06	The system performed certain actions I was not aware of.
07	Critical / relevant information for me as a user was readily available at all times.
08	The system gave specific actions or choices (visual) precedence over others.
09	The system required me to perform specific (unwanted) actions in order to proceed to the next step in a process.
10	The system guided my behaviour in a way that benefited the designer of the system (e.g., the online company) in the end.
11	I think that within this system, the user is put at the centre of attention.
12	I think that this system brings harm to its users.
IT13	Using the system, I felt that I had control over my own actions and choices.
14	The system performed certain actions without my consent.
15	The system pushed me into spending more money than I originally anticipated.
16	I felt I had control over the sharing of my personal information.
17	I felt the system used my emotions to trick me into performing certain actions.
18	The system caused me to spend unnecessary time, energy and attention to perform a desired action or select a desired choice.
19	The actions I performed using the system always resulted in the expected outcomes.
20	The wording used in the system was explicit and clear.
21	The system took good care of my individual welfare.
22	I felt deceived/misled by the system.
23	I felt the designer of the system had the user's best interests in mind.
24	I felt like the system helped me in making better choices, that would benefit me as a user in the end.
25	The possible range of actions I could perform within the system was clear to me at all times.

3.3.4 The SDS Questionnaire Study

To select the final items for the SDS out of the pool of all potential items, a study needed to be conducted.

Study Task

Participants were asked to fill out a questionnaire containing all the potential questionnaire statements. They did this for either the "really dark" or "really bright" system -depending on which system they interacted with in Study 1. Participants were told to respond to the statements on a 5-point Likert Scale, with (1) representing 'totally agree' and (5) representing 'Totally Disagree'.

Figure 26 provides an example of one of the statements that participants had to respond to.

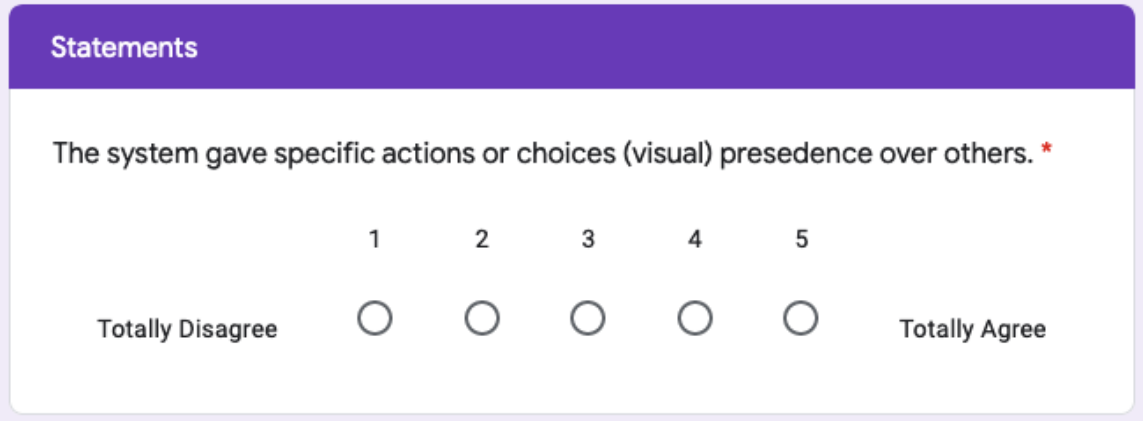


Figure 26: One of the Statements of the SDS Questionnaire.

Study Design

The design of the study was decided to be a between-subjects design, as this matched the design of Study 1 better. Additionally, having participants interact with- and fill out questionnaires for both systems would have doubled the time needed to complete Study 2. As we wanted to use the same participants for all three study objectives, and keep them motivated to reach the end of the whole overarching study, this was not desired.

Figure 27 provides some insight into how Study 1 and Study 2 were related to each other.

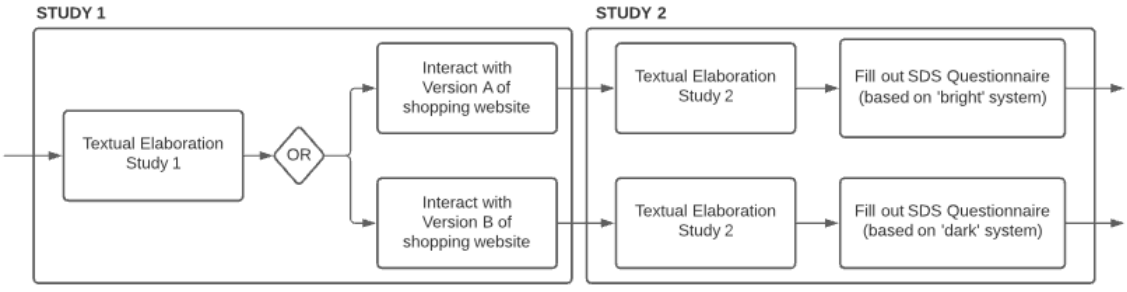


Figure 27: Flow of Study: Detailed Relation Study 1 & Study 2

Procedure

After participants finished Study 1, they were presented with a web page containing the textual elaboration upon Study 2. Participants were told that they had to respond to a set of statements on a 5-point Likert scale, while keeping in mind the web shop they had just interacted with. They were also told not to include the experimental UX pop-ups (Study 1) into their ratings. A button at the bottom of the web brought participants to the Google Forms environment, and presented them with the SDS questionnaire.

The confirmation page of the Google Forms questionnaire provided participants with a link which lead them back to the online experimental environment. When clicking this link, participants found themselves on the explanation page of Study 3. This page told participants that they had finished the second study, and provided them with a textual elaboration on the next (and last) one.

3.3.5 Selecting the Final Items: Data Preparation

Which items will be selected for the final SDS questionnaire will be based upon the analysis of the data that was gathered in the SDS Study. In order to work with this data, it needed to be prepared and cleaned.

Collected data from the SDS study was automatically saved to Google Sheets. The data sheets of both the "dark" and "bright" system were converted to a .csv file and loaded into the statistics tool R.

A data frame was created which combined the sheets of both systems. Within this data frame, each row represented a participant. Each column in the data frame represented a questionnaire item. The first column of the data frame represented the Version the participant was in, with A being the "really bright system", and B being the "really dark system". Apart from the cells in the first column, each of the cells in the data frame contained a likert scale value (1-5).

The Results section (Section 4.3) will reveal which Questionnaire Items will be selected for the final SDS in order to validly capture the ‘darkness’ of a certain system.

3.3.6 Scoring the SDS

However, selecting a set of questionnaire items is not the only thing that needs to be done in order to develop the System Darkness Scale. As scores for individual questionnaire items are not meaningful on their own, some formula needs to be decided upon to end up with a *single* SDS ‘darkness’ score. "How to score the SDS" strongly depends on the final number of questionnaire items, and will therefore be discussed in the Results section as well.

3.4 Study 3: Assigning a Severity Score to Dark Patterns

The last aim of the current study is to assign a severity (or Darkness) score to each of the 17 Dark Pattern types (as presented in table 3), and rank the Dark Patterns accordingly. As such, the main question of Study 3 can be regarded as follows: *"Which Dark Patterns do users perceive to be the most severe?"*

3.4.1 Study Tasks

In order to answer this question, participants had to fill out a questionnaire. Within this questionnaire, participants were presented with 17 situations in which Dark Patterns were involved. The situations were described in text and supplemented with an example image. Participants were asked to imagine themselves being in the situation at hand, and indicate how they felt about the situation. They could do this on a 5-point Likert scale, ranging from (1) "I don't mind it" to (5) "I find it deeply annoying". Additionally, participants were asked to indicate how frequently they had experienced a similar situation before.

Figure 28 presents one of the 17 Dark Pattern situations, and the two questions participants had to answer based on the situation.

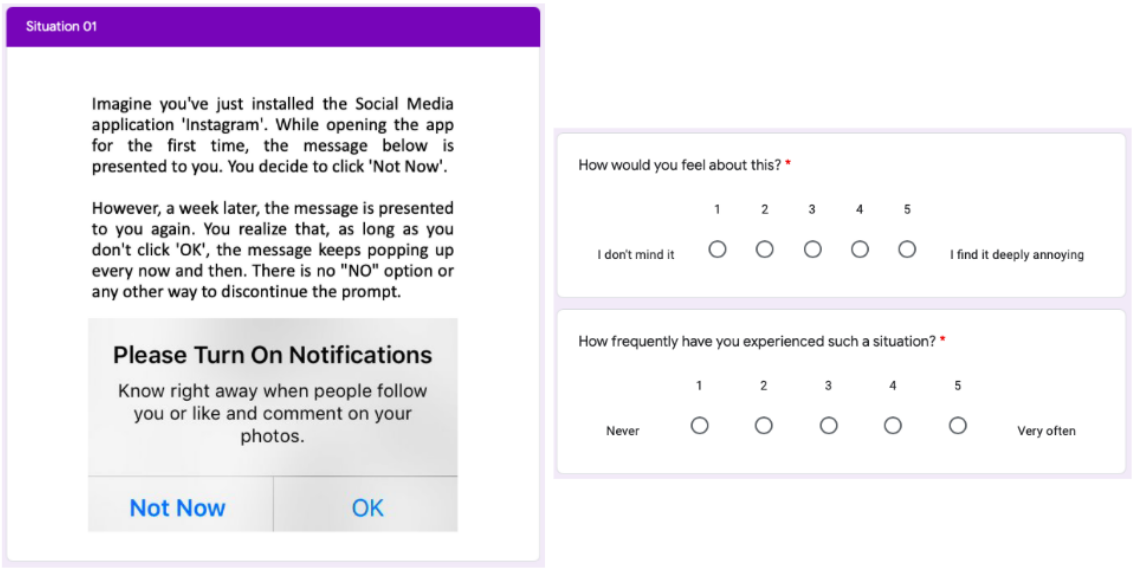


Figure 28: Example of one of the Situations in the Severity Questionnaire.

3.4.2 Study Design

Different from Study 1 and 2, the design of Study 3 was not a between-subject design. As the aim of Study 3 was not directly related with finding out whether there is a difference between condition x and y, using a between-subject design made no sense. Study 3, therefore, treated all participants (whether they performed in version A or B of the previous studies) as one group.

Figure 29 provides some insight into how Study 2 and Study 3 were related to each other.

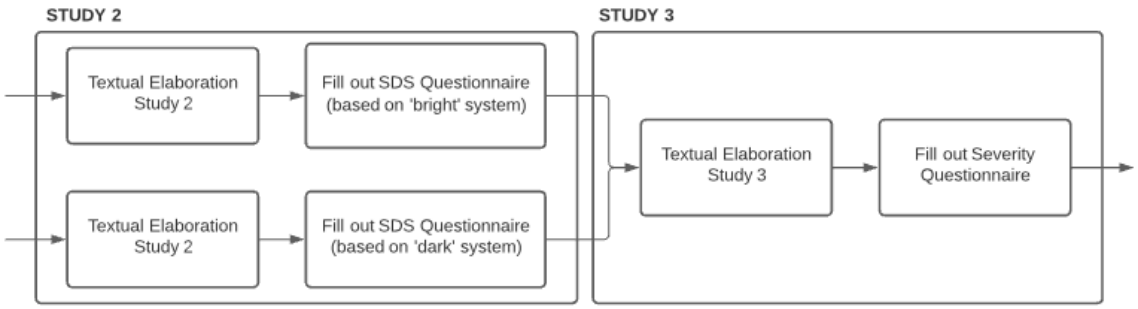


Figure 29: Flow of Study: Detailed Relation Study 2 & Study 3

3.4.3 Research Material

The situations in which Dark Patterns were involved were created by the researcher herself. First, some visual examples of Dark Patterns were collected. Most of them were derived from the website https://www.darkpatterns.uxp2.com. Based on the visual examples, the researcher made up some stories. All of the stories described a certain interaction, in which a Dark Pattern was encountered.

Within the descriptions of the situations, Dark Patterns were not specifically mentioned by their name, nor were they specifically explained with their definition. Additionally, the situations were described in a neutral manner. This way, the researcher hoped not to prime participants into thinking a situation was bad before they filled out the questions. By keeping descriptions as neutral as possible, the *actual* feelings of the participant could be captured by the questionnaire.

The different situations related to each of the 17 Dark Patterns were checked for clarity by two potential participants. Slight changes were made according to their comments, most of them being grammar mistakes. The 17 final situations can be found in Appendix A.6.

3.4.4 Procedure

After participants finished Study 2, they were presented with a web page containing the textual elaboration upon Study 3. At the bottom of the page, a button was presented. Clicking this button resulted in participants being redirected to the Google Forms environment, in which they could find the questionnaire. Within the questionnaire, participants were presented with one situation at a time. After filling out their responses on the two questions for a certain situation, the next situation was presented to them.

The study ended when participants had indicated ratings for all of the 17 situations. As soon as this happened, participants were redirected back to the online experimental environment and were shown the final web page of the study, which included the concluding remarks.

3.4.5 Variables

The variable of interest within Study 3 was 'Level of Severity'. As this variable was deemed to be a bit ambiguous in and of itself, the researcher decided to not use the term 'severe' in the questionnaire of the study. In order to still capture the 'Level of Severity', while at the same time using a scale that is more intuitive for participants to use, the researcher decided to use "I find it deeply annoying" and "I don't mind it" as the opposite ends of the 5-point Likert Scale.

Another variable that was part of Study 3 was 'Frequency of Encounter'. This variable was *not* the variable of interest, but was collected *only* to provide some additional insights. Reason for this was that the researchers could imagine there being a relationship between the frequency of encounter and indicated level of severity. For example, when a Dark Pattern is more frequently encountered, people might get 'used to it' and therefore indicate it to be less severe. 'Frequency of Encounter' was recorded on a 5-point Likert Scale, ranging from (1) 'Never', to (5) 'Very Often'.

3.4.6 Data Preparation

Collected data from the DPDS study was automatically saved to Google Sheets. The data sheet was converted to a .csv file and loaded into the statistics tool R.

A data frame was created in which each row represented a participant. Sets of two columns represented a Dark Pattern Situation, with the first column containing the likert scale value of the 'severity' question (1-5), and the second column containing the likert scale value of the 'frequency' question (1-5).

4 Results

4.1 Descriptive Statistics

Over the course of 3 weeks 92 participants (n=92) were selected to partake in the study. All participants performed all three study parts. For Study 1 and Study 2, 46 participants (50%) performed in Version A and 46 participants (50%) performed in version B. The participants in version A and B have very similar age averages. Additionally, the absolute amount of males and females in each version is fairly consistent. A full overview of participant descriptives is provided in table 7.

Table 7: Descriptive Statistics of Participants: Sample Size, Age & Gender.

	Sample Size	**Age**	**Gender**
Version A (Control Condition)	46	min = 15, max = 63, mean = 30.3, median = 24.5	Male = 19, Female = 27, Non-binary = 0
Version B (Dark Patterns)	46	min = 19, max = 64, mean = 27.5, median 24	Male = 20, Female = 25, Non-binary = 1
Total	92	min = 15, max = 64, mean = 28.9, median = 24	Male = 39, Female = 52, Non-binary = 1

Table 8 presents an overview of the educational background of all participants. The absolute amount of participants with the same educational background is fairly consistent for Version A and B, for all educational levels. None of the participants indicated to have obtained a professional degree (PhD).

Table 8: Descriptive Statistics of Participants: Level of Education.

	Primary School	**High School**	**MBO**	**HBO**	**Bachelor**	**Master**
Version A (Control Condition)	0	13	2	4	18	9
Version B (Dark Patterns)	1	7	2	6	20	10
Total	1	20	4	10	38	19

4.2 Study 1: Performing an Experiment on User Experience

This section provides an overview of all the data that was gathered during the User Experience experiment, and the calculations that were made for answering the research question: "Does the use of Dark Pattern strategies in UI design influence a user's experience with
a service or product in comparison to these malicious strategies being absent?".

In order to answer this research question, some hypothesis tests needed to be performed. In what follows, the test results of both the hypotheses introduced in section 3.2.6 will be provided. The null hypotheses formulated within this section were to be tested against a confidence interval of 95%, thus giving a significance level of $\alpha = 0.05$. This is an acceptable level given the limited number of participants, but strict enough to provide reasonable validity to the results.

4.2.1 H1: Level of Satisfaction

This subsection provides the results for the first hypothesis:

$H1$ The use of Dark Pattern strategies in the UI design of a web page (Condition B) leads to a decrease in the Level of Satisfaction relative to the UI design of a web page in which malicious design strategies are absent (Condition A).

For hypothesis one, the null and alternative hypothesis were formulated as follows:

$H1_0$ Level of Satisfaction 'page X' in Version A = Level of Satisfaction 'page X' in Version B

$H1_a$ Level of Satisfaction 'page X' in Version A > Level of Satisfaction 'page X' in Version B

Within the null and alternative hypothesis, 'page X' represents one of the seven shopping task-related web pages (e.g., home, postcard categories, postcard overview, postcard details, shopping cart, check-out, and place order). For each of the seven pages, a statistical test needed to be performed.

Statistical Test

Testing whether the Level of Satisfaction of participants is higher when no Dark Patterns are present in the interface of a web page (Version A) compared to when Dark Patterns are present in the interface of a web page (Version B) would ideally be done through a one-tailed independent t-test.

One assumption of the parametric independent t-test, however, is that the sample data has a normal distribution. As such, a Shapiro-Wilk Normality Test was performed on the 'Level of Satisfaction' data for each of the seven shopping task-related web pages. The test showed that none of the data met the normality assumption. Because of this violation, the Mann Whitney U-test was selected as the most applicable statistical test, as this test is often considered the non-parametric alternative to the independent t-test (Field, 2013).

Table 9: Average 'Level of Satisfaction' ratings for both Version A (*no* Dark Patterns present) and Version B (Dark Patterns present) of the experiment.

	Version A	Version B
Home	3.6	3.7
Categories	3.7	3.8
Overview	3.9	3.8
Details	3.7	3.9
Shopping Cart	3.8	2.7
Checkout	3.8	4.0
Place Order	3.9	2.4

Table 10: Results of the Mann Whitney U-test. If the p-value of the test is below 0.05, the Level of Satisfaction scores for version A and B statistically differ.

	U	z	p
Home	952	-.915	.184
Categories	973	-.717	.234
Overview	987	-.598	.285
Details	961	-.817	.210
Shopping Cart	497	-4.605	.000*
Checkout	985	-.617	.270
Place Order	355	-5.654	.000*

Hypothesis Testing

Table 9 shows the average 'Level of Satisfaction' ratings that were provided for Version A and B of the experiment. The results of the one-tailed Mann Whitney U-tests are provided in table 10.

The results show that:

- For the *'Shopping Cart' page*, Level of Satisfaction scores of participants participating in Version A of the experiment (no Dark Patterns in UI) were higher than those of participants in Version B of the experiment (Dark Patterns in UI). The Mann Whitney U-test indicated that this difference was statistically significant: U ($N_{VersionA}$ = 46, $N_{VersionB}$ = 46) = 497, z = -4.605, p = .000.
- For the *'Place Order' page*, Level of Satisfaction scores of participants participating in Version A of the experiment (no Dark Patterns in UI) were higher than those of participants in Version B of the experiment (Dark Patterns in UI). The Mann Whitney U-test indicated that this difference was statistically significant: U ($N_{VersionA}$ = 46, $N_{VersionB}$ = 46) = 355, z = -5.654, p = .000.
- For all other pages, no significant difference was found.

As such, for the pages 'Shopping Cart' and 'Place Order' we can reject the null hypothesis and accept the alternative hypothesis. For all the other pages, the null hypothesis can not be rejected.

Visualisations

Figure 30, 31, 32 and 33 substantiate the results of testing hypothesis one. As can be seen, the histograms of the 'Shopping Cart' and 'Place Order' page (figure 30 and 31) show a clear difference in means (dotted line) for Version A (red) and B (blue) of the shopping website.

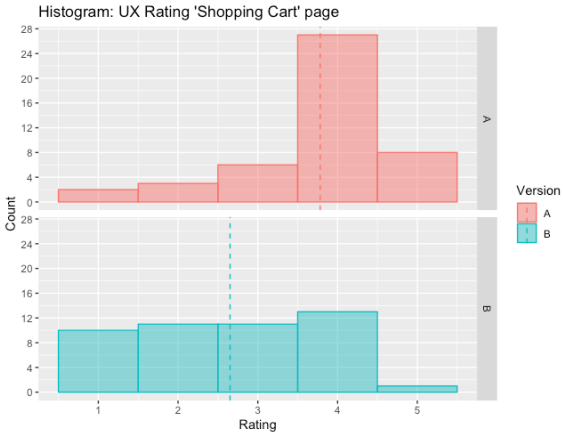


Figure 30: Level of Satisfaction Scores participants provided for the 'Shopping Cart' page.

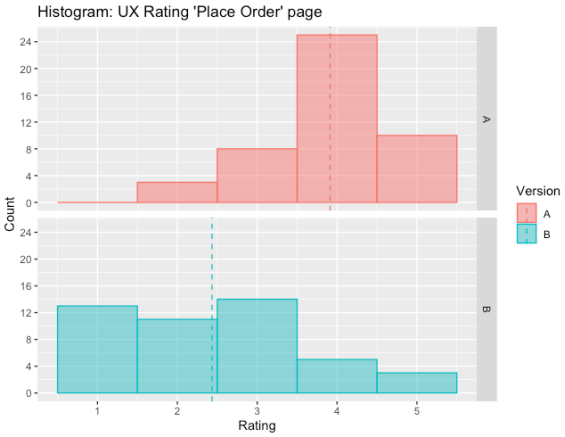


Figure 31: Level of Satisfaction Scores participants provided for the 'Place Order' page.

In contrast, the means for Version A and B of the non-significant pages are much closer to each other -they are almost the same. In figure 32 and 33, the plots of the 'Categories' and 'Details' page provide an insight into what the data of the non-significant pages looked like. The resemblance in means between version A and B is not surprising when looking at the shape of the histograms. For both web pages, the shape of the histograms of Version A and B almost 100% overlap. This is also the case for the 'Home', 'Overview' and 'Checkout' page.

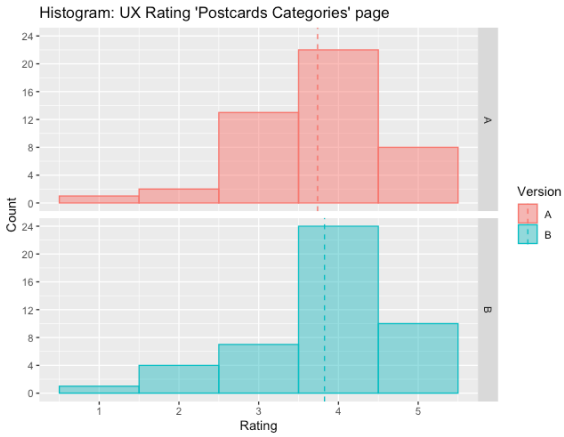


Figure 32: Level of Satisfaction Scores participants provided for the 'Categories' page.

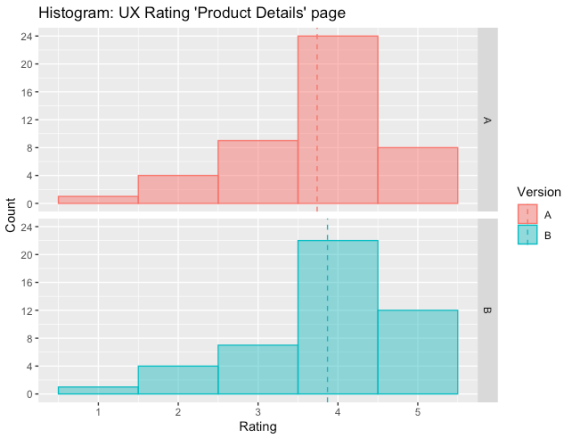


Figure 33: Level of Satisfaction Scores participants provided for the 'Product Details' page.

The results of the various Mann Whitney U-tests are also clearly visible in the Customer Journey Map (figure 34), which was created based upon the average 'Level of Satisfaction' ratings for each of the pages within Version A and B of the experiment.

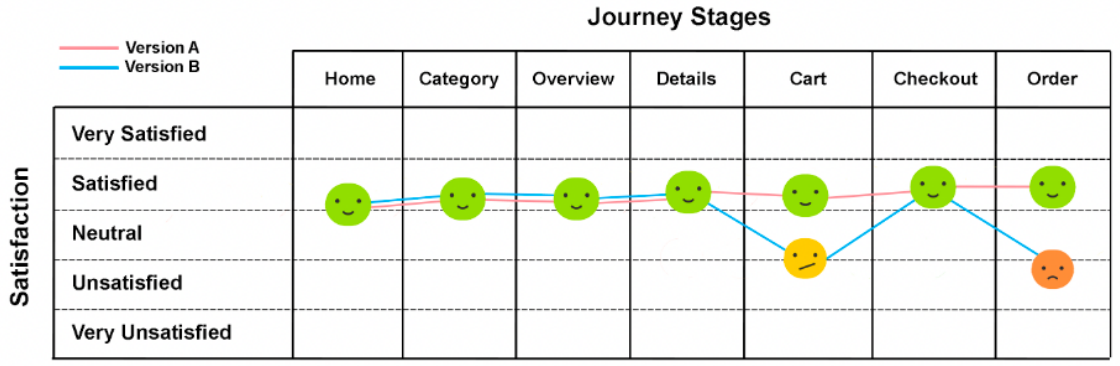


Figure 34: Customer Journey Map of Version A (pink) and B (blue).

4.2.2 H2: Recommendation Score

This subsection provides the results for the second hypothesis:

H2 The use of Dark Pattern strategies in the UI design of a website (Condition B) leads to a decrease in the Recommendation Score relative to the UI design of a website in which malicious design strategies are absent (Condition A).

For hypothesis two, the null and alternative hypothesis were formulated as follows:

$H2_0$ Recommendation Score Version A = Recommendation Score Version B

$H2_a$ Recommendation Score Version A > Recommendation Score Version B

In contrast to hypothesis one, hypothesis two is related to measuring the User Experience for the whole website, rather than for specific web pages. As such, only one statistical test needed to be performed.

Statistical Test

Testing whether the Recommendation Score of participants is higher when no Dark Patterns are present in a website (Version A) compared to when Dark Patterns are present in a website (Version B) would ideally be done through a one-tailed independent t-test.

However, similar to the data of the UX pop-up, the Referral pop-up data did not meet the normality assumption of the independent t-test. As such, the Mann Whitney U-test was once more selected as the most applicable statistical test.

Hypothesis Testing

Results of the one-tailed Mann Whitney U-test showed that the Recommendation Scores of participants participating in Version A of the experiment (no Dark Patterns present) were indeed higher than those of participants in Version B of the experiment (Dark Patterns present). The Mann Whitney U-test indicated that this difference was statistically significant: U ($N_{VersionA}$ = 46, $N_{VersionB}$ = 46) = 544, z = -4.068, p = .000. This results in the rejection of the null hypothesis and the acceptance of the alternative hypothesis.

Visualisations

Figure 35 and figure 36 visualise the data that was obtained through the Referral Rating pop-ups. In both the figures, one can clearly see the differences in the Recommendation Scores that participants provided for either Version A and B of the experiment.

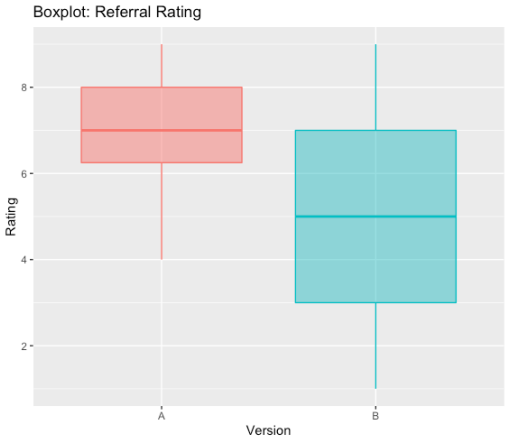


Figure 35: Boxplot of Referral Rating Data.

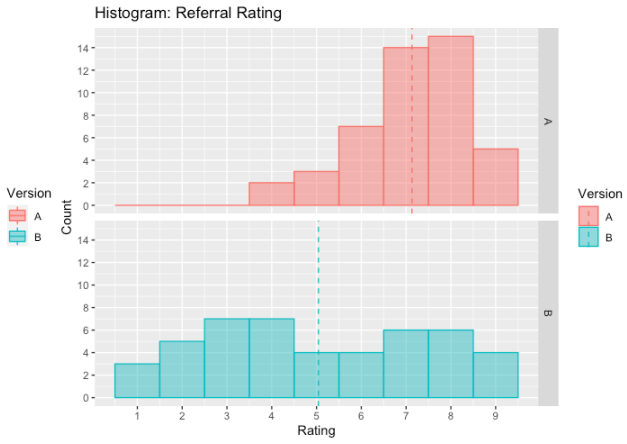


Figure 36: Histogram of Referral Rating Data.

In the boxplot in figure 35, the y-axis represents the Recommendation Score, whereas the x-axis represents the version (with A being the shopping website *without* Dark Patterns, and B being the shopping website *including* Dark Patterns). The boxplot of Version A shows that the median Recommendation Score provided by participants is 7. Most participants scored the shopping website between a 6 and an 8, but some higher and lower ratings were provided as well. However, none of the participants rated the web shop lower than a 4. The median score for version B is much lower: 5. The spread of boxplot B also greatly differs from boxplot A. For Version B, a greater variety of Recommendation Scores was provided by participants, from '1's to '9's.

The histogram in figure 36 shows this difference in spread as well. Within this figure, the x-axis represents each of the Recommendation Score options (1-10), whereas the y-axis represents the number of participants that filled out each option (frequency). In Version A, the higher Recommendation Scores have the highest frequency, whereas in Version B, the frequency of the various response options is more evenly distributed, with slightly higher frequencies for the lower scores. The mean Recommendation Score (dotted line) greatly differs for Version A and B of the web shop as well.

4.2.3 Additional insights: Recommendation Scores Explained

As mentioned in section 3.2.1, in the Referral pop-up, participants were provided with the opportunity to explain their recommendation score. By doing this, the researchers hoped to gain some more insights into the likes and dislikes of the (interface design of the) web shop as well.

Although leaving an explanation upon the recommendation score was not required, out of all 92 participants, only 14 participants did *not* explain their score. The other 78 participants did. This resulted in the acquisition of 40 explanations for Version B and 38 explanations for Version A.

Explanations Version B

Looking at the explanations of Version B, there are a few 'themes' or 'aspects' that really stand out. Although not primed with (the various types of) Dark Patterns, participants mentioned quite a few of them in their explanations. It is important to mention that Dark Patterns were often not referred to by their specific 'name', but were rather described in some way.

The Dark Pattern that was the most often mentioned was "Hidden Costs". This Dark Pattern was displayed on the 'Place Order' page -at the very end of the checkout process. Out of all 40 explanations for Version B, the "Hidden Costs" Dark Pattern was referred to 23 times. For example, P39 stated:

> P39 *"Seeing the large amount of tax after proceeding with my order is not user-friendly. When I place a product in the shopping cart I expect it to be the same price while proceeding. When this happens in other web shops, most of the time I cancel my order because it seems like I am getting screwed (that the vendor is lying to me). It*

is like picking up a product in a physical store that has a price tag of 5 euros, and when you go to the cashier it says 10 euros."

In addition to the "Hidden Costs" Dark Pattern, the "Sneak into Basket" Dark Pattern (displayed on the 'Shopping Cart' page) was often referred to as well. This Dark Pattern was mentioned in 11 of the 40 explanations. For example, P19 stated:

*P*19 *"I did not like the fact that stamps were automatically added to my shopping basket, as I did not ask for them."*

Other Dark Patterns that participants referred to were "Toying with Emotions" (6x), "Hidden Information" (7x), "Trick Questions" (6x), "Preselection" (5x) and "Price Comparison Prevention" (1x). The following quotations provide some idea of *how* they were referred to:

*P*29 *"The site felt a bit pushy with the countdown timer counting down to zero, and showing how many people are currently looking at certain products."*

*P*07 *"The checkbox for promotional emails was hard to find, already checked (I thought we had an opt-in policy?), and the sentence was hard to understand."*

*P*32 *"I didn't like that you had to calculate the price per card yourself."*

Next to Dark Patterns, some functionality issues of the website itself were mentioned. These included 'not receiving the offer in the inbox' and the fact that 'there was no product filtering option'. Positive aspects mentioned by participants which were assigned to Version B were that the shopping website was 'well structured' and 'good looking'.

Explanations Version A

Whereas the explanations of participants in Version B mostly stressed negative aspects of the shopping website, the opposite was true for the explanations of participants in Version A. Participants praised the design and navigation of the web shop. They also mentioned that they found the web shop 'user friendly', 'structured' and 'inviting'.

The few negative aspects that were mentioned were all related to the functionality of the web shop (rather than the design of the User Interface). Some participants mentioned that they would have liked the option to edit products from the 'shopping cart' page. Others stated that they missed a 'product filtering' option or did not receive the promotional offer in their inbox.

4.3 Study 2: Creating a 'System Darkness Scale'

This section provides an overview of all the calculations that were made in order to select a number of questionnaire items out of a potential list of 25 of them. The final set of questionnaire items should -all together- be measuring the *darkness* of a certain system.

The final System Darkness Scale (SDS) is presented at the end of this section. A short description upon how to score the SDS questionnaire is provided here as well.

4.3.1 Selecting Items for the SDS

In making judgements about which items should be selected for the final System Darkness Scale, the current study followed the uni-dimensional scaling approach of Trochim (2021). Several analyses were performed, each of them leading to the *elimination* of a set of potential questionnaire items. The final set of *selected* questionnaire items is presented at the end of this subsection.

Calculating Item-Total Correlation

Within the current study, we want to construct a single 'darkness' score for a particular system, to be able to''compare' the system to other systems. In order for the final System Darkness Scale (SDS) to produce a reliable single 'darkness' score, it is important that the final items of which the SDS consists are all measuring this construct of 'darkness', and hence produce responses that are in agreement with each other. *Internal consistency* of the SDS, therefore, is of great importance.

Reliability analyses are often performed to test internal consistency. Within these analyses, the current reliability (Cronbach's alpha) of a multi-item scale is measured. Researchers determined a reliability score -or Cronbach's alpha- of at least 0.7, but more preferably higher than 0.8, to be desirable. If the score is lower than 0.7, one could improve it by eliminating certain questionnaire items.

The Cronbach's alpha of measuring the construct of 'darkness' by using all 25 potential questionnaire items turned out to be very low: Cronbach's alpha = 0.5. As such, it became clear that a number of questionnaire items had to be eliminated.

The item-total correlation is a test which can help in determining *which* items are inconsistent in measuring the construct at hand, and therefore should be eliminated. The test calculates the Pearson correlation coefficient for pairs of scores, where one item of each pair is an item score, and the other item is the summed score of all items. The greater the value of the coefficient, the stronger the correlation, and the better the particular item contributes to measuring the construct at hand. Items with stronger correlations should therefore be retained. On the contrary, items with weaker correlations should be eliminated. Eliminating these items would lead to improving Cronbach's alpha, and therefore the internal consistency of the System Darkness Scale.

As such, the goal of the first analysis was to eliminate the potential questionnaire items that have a low correlation with the (total) summed score. Table 11a shows the correlations between each of the items and the total summed score.

Table 11: Results of two of the analysis that were performed. Items with high item-total correlation and high discrimination (e.g., high t-values) are potential candidates for the final SDS (green coloured items). The red coloured items were selected as potential candidates for the SDS by the item-total analysis (as they have high correlation coefficients), but eliminated as a result of running the discrimination analysis (as their t-values are rather low).

(a) Item-total correlation for each of the 25 potential SDS items (ordered from high to low).

	Item	Correlation
1	10	0.630
2	14	0.608
3	06	0.587
4	01	0.558
5	15	0.543
6	09	0.542
7	17	0.468
8	22	0.468
9	08	0.447
10	12	0.405
11	02	0.387
12	04	0.379
13	18	0.367
14	20	0.085
15	03	0.084
16	24	0.078
17	05	0.065
18	11	0.059
19	21	0.052
20	19	-0.003
21	07	-0.022
22	25	-0.048
23	16	-0.100
24	23	-0.144
25	13	-0.167

(b) T-values of t-tests for each of the 25 potential SDS items (ordered from high to low).

	Item	T-value
1	15	-
2	14	65.1
3	06	43.0
4	22	39.9
5	01	33.7
6	18	33.7
7	17	33.0
8	19	32.6
9	09	26.4
10	07	22.4
11	02	22.2
12	03	20.9
13	25	20.9
14	21	20.7
15	08	20.5
16	04	20.2
17	20	19.9
18	23	19.7
19	24	19.3
20	11	18.8
21	12	17.4
22	05	17.3
23	13	17.2
24	16	16.2
25	10	14.8

Unfortunately, no fixed rules exists on *when* to eliminate an item. The researcher of the current study decided to eliminate all items with a correlation with the summed score of less than 0.4. As such, the top 10 items represented in table 11a remained possible candidates for inclusion in the final SDS.

When running the reliability analysis with the top 10 items, Cronbach's alpha -and therefore the internal consistency of the SDS- improved drastically: Cronbach's alpa = 0.9.

Calculating Discrimination

In addition to total-item correlation, item discrimination was calculated as well. Item discrimination analysis checks whether a difference exists between two sample sets -in our case 'participants in Version A' and 'participants in version B'. As we want to select the items that lead to extreme opposite responses for Version A and B, we want the item discrimination to be as high as possible.

The goal of the second analysis, therefore, was to eliminate all items with low discrimination. For each of the items, the data of the quarters of participants that gave the highest and lowest ratings were selected. In other words, out of all 92 participants (100%), for each item, the data of the 23 participants (25%) that gave the highest ratings were selected, as well as the data of the 23 participants (25%) that gave the lowest ratings. Average ratings were calculated for both these groups. Thereafter, t-tests were performed on the difference between the average value for each item's top and bottom quarter participants.

Table 11b shows the t-values resulting from the t-tests. Higher t-values mean that there is a greater difference between the highest and lowest participants. In other words, items with higher t-values are better discriminators. No t-value is shown for IT15, as performing a t-test is not possible on 'perfect' data (e.g., the data of the participants in the top quarter consisted of only '5' and the data of the participants in the bottom quarter consisted of only '1').

As we want our final SDS items to have *high* item-total correlation and *high* discrimination (e.g., high t-values), the researcher took the top 10 items that were selected as possible candidates for the SDS in the previous analysis, and took a close look at their t-test values within the current analysis.

There were several items that were represented in the top 10 of both the Correlation- and T-value table (table 11): Item 01, 06, 09, 14, 15, 17 and 22 (coloured green). These items were selected to proceed to the last analysis. Item 10, 08 and 12 (coloured red) were eliminated. Although these items received high correlation scores, they all scored relatively low when looking at their T-values, with item 10 and item 12 even ending up within the bottom 5 of the T-value ranking.

Calculating Item Inter-correlations

As the System Usability Scale (SUS) reported to have very close inter-correlations between all of the selected items (± 0.7 to ± 0.9), inter-correlations were checked for the pre-selected items of the SDS as well. Table 12 shows these inter-correlations.

Table 12: Inter-correlations between all of the pre-selected items. For inter-correlations to reach an adequate level, item 09 and 17 were eliminated from the list of potential candidates for the SDS (red coloured cells).

	01	**06**	**09**	**14**	**15**	**17**	**22**
01	x	0.5	0.5	0.6	0.6	0.3	0.6
06	0.5	x	0.5	0.7	0.6	0.2	0.5
09	0.5	0.5	x	0.5	0.4	0.3	0.4
14	0.6	0.7	0.5	x	0.7	0.3	0.7
15	0.6	0.6	0.4	0.7	x	0.4	0.7
17	0.3	0.2	0.3	0.3	0.4	x	0.4
22	0.6	0.5	0.4	0.7	0.7	0.4	x

Table 12 shows that the inter-correlations between the former selected SDS items are lower than the inter-correlations of the items of the SUS. To make sure the inter-correlations of the SDS reach an adequate level, the researcher decided to eliminate another two items from the list: 09 and 17. By doing this, the inter-correlations between all of the selected items moved from (± 0.2 to ± 0.7) to (± 0.5 to ± 0.7).

Comparing Test Results & Visualisations

Before performing any statistical tests, the researcher created visualisations for each of the potential SDS items. The visualisations provided the researcher with a first insight into the response frequency for each of the items. As already mentioned multiple times, questionnaire items with extreme opposite responses for the different systems (A and B) would be potential candidates for the final SDS.

Based on the visualisations, the researcher expected items 06, 14, 15 and 22 to be included in the final SDS. To provide an example of what the visualisations of these items looked like, figure 37 shows the plot of one of them. Within this plot, you can clearly see the extreme opposite responses.

One of the items the researcher ranked as a definite 'do not include' (based on the visualisations) was item 16. Figure 39 shows that for this item, the spread of the plot almost 100% overlaps for both the 'Dark' and 'Bright' version of the system. This is not desired, as we want *high* discrimination.

Before running the tests, the researcher categorised (amongst others) item 01, 09 and 17 as 'doubtful cases'. As can be seen in figure 38, for those items, one cannot state that there is an overlapping spread, however the spreads can not be regarded extreme opposites of each other as well.

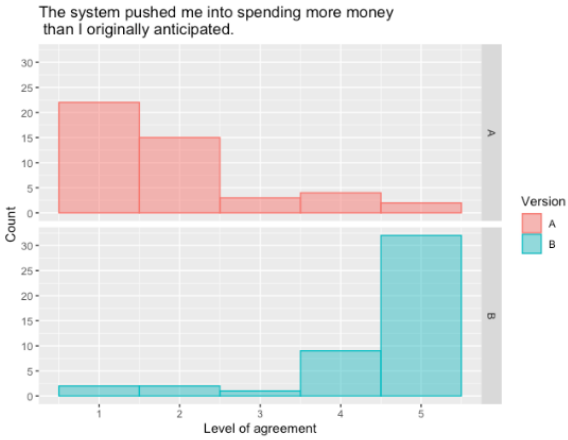


Figure 37: Plot of item 15: Potential candidate for SDS.

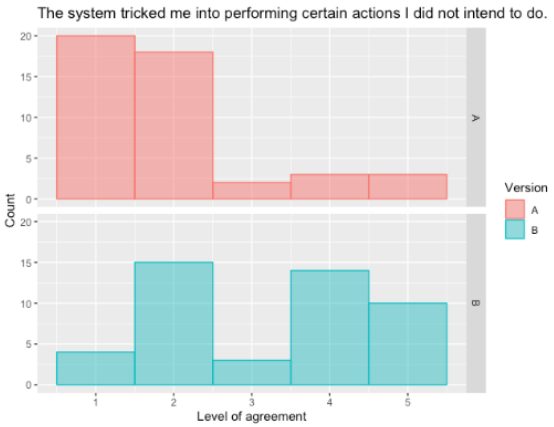


Figure 38: Plot of item 01: Doubtful case.

Figure 39: Plot of item 16: Expected to be eliminated.

We can state that the expectations that were made based on the visualisations quite accurately match the results of the (statistical) analyses:

- Item 06, 14, 15 and 22 were indeed selected as items for the final SDS questionnaire.
- Item 16, indeed, ended up at the bottom of both table 11a and 11b, and was eliminated as such.
- Item 01, 09 and 17 indeed ended up as 'doubtful cases'. They worked their way through several analyses, but only one of them (item 01) made it to the final SDS questionnaire.

The Final SDS

Figure 40 presents the final System Darkness Scale.

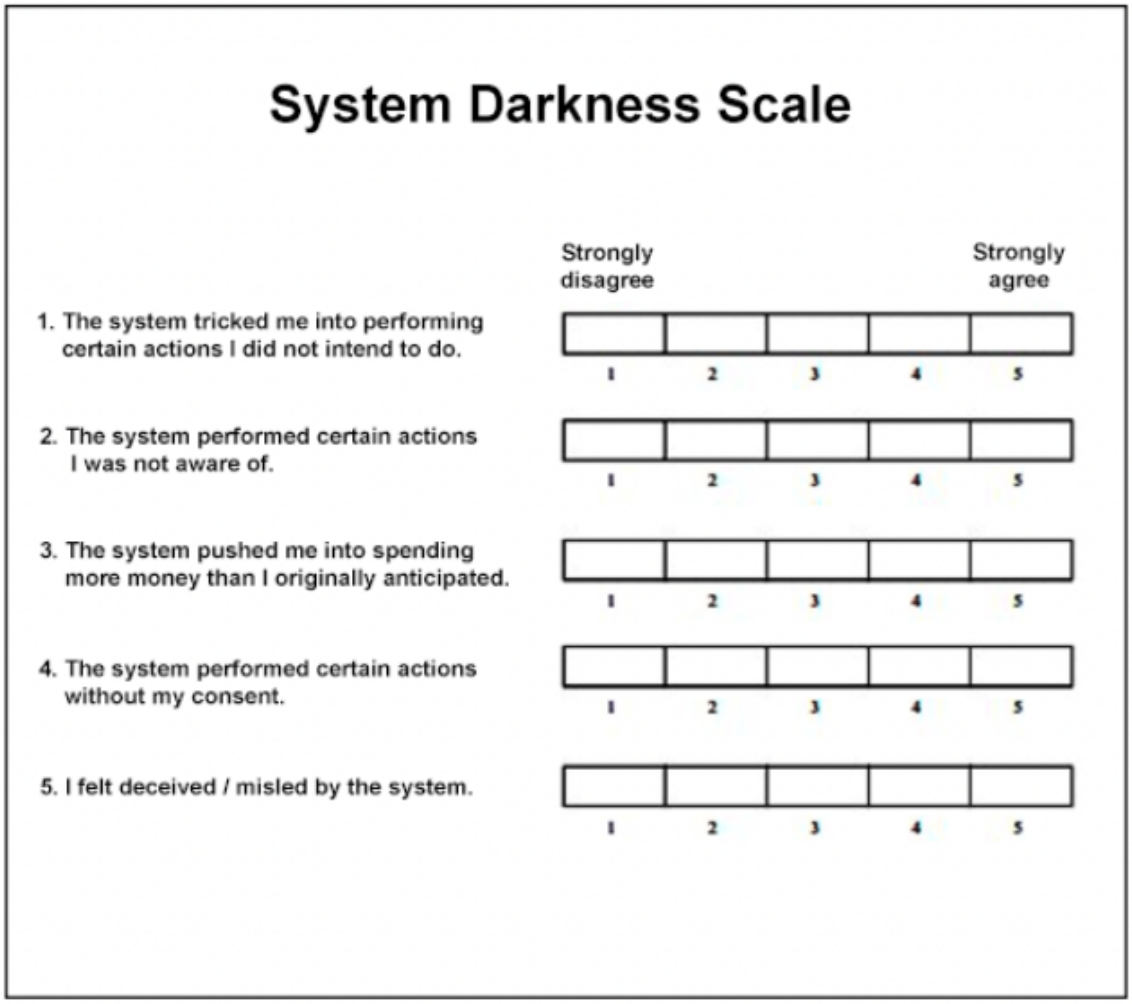


Figure 40: Final System Darkness Scale (SDS)

4.3.2 Scoring the SDS

A Likert scale is a composite of multiple Likert items. The System Darkness Scale consists of five such items. Likert scales are 'summated' scales, so called because a respondent's answers on each question item are summed to give their overall score on the attitude or value of interest -in this case 'Darkness'. In order to retrieve this 'overall score', some calculation needs to be decided upon.

In the SDS, there are 5 response options for each of the 5 items. The researcher decided to code 'Strongly Disagree' as a 0, and 'Strongly Agree' as a 4. In order to retrieve the overall darkness score, the respondent fills out all questionnaire items, and then sums up the scores related to the chosen response options. Once the scores are summed, they can be 'rescaled' into a more intuitive range. For example, the sum of scores on the SDS ranges between 0 (all 0s) and 20 (all 4s). By multiplying the sum of scores by 5, summed scores are converted into an 'overall SDS score' that ranges between 0 and 100.

Within this range, a score of 0 represents a 'really bright (or even harmless) system', whereas a score of 100 represents a 'really dark (severe) system'.

Figure 41 gives an example of a scored SDS scale.

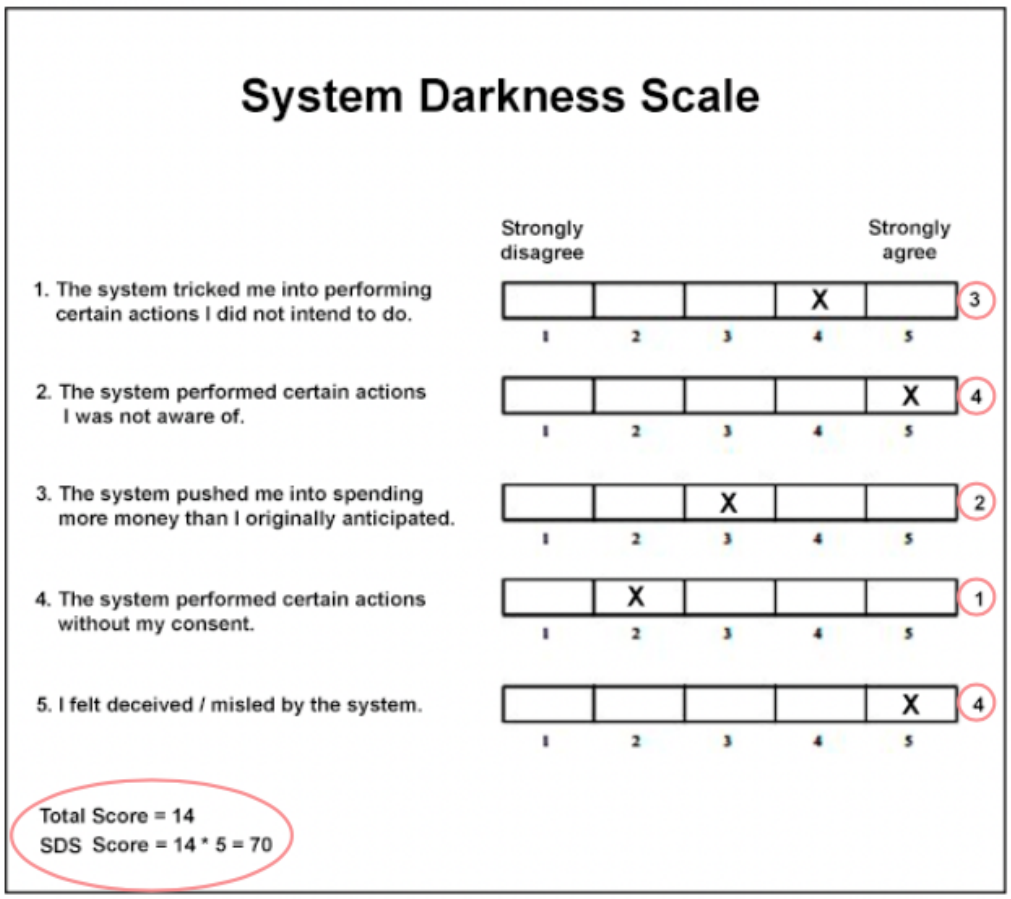


Figure 41: Scoring the SDS: example.

4.4 Study 3: Assigning a Severity Score to Dark Patterns

This section provides an overview of all the calculations that were made for assigning severity scores to each of the 17 Dark Pattern types, and categorising them into severity-impact categories. The last part of the section provides some more insights into how frequently participants encountered the various Dark Pattern types, and whether this could have had an effect on the perceived severity of a Pattern.

4.4.1 Scoring Dark Pattern Types

As already mentioned before, for each of the situations (or: Dark Pattern types) presented in the DPDS questionnaire, participants were asked to indicate how they would feel if they found themselves in the situation. Google Forms automatically coded the response 'I don't mind it' as a 1, and 'I find it deeply annoying' as a 5. However, to be able to create a more intuitive and comprehensive rating in the end, the researcher decided to translate the 'I don't mind it' ratings from 1s to 0s, and 'I find it deeply annoying' ratings from 5s to 4s.

After this translation was finished, for each of the situations (Dark Pattern types) the average rating was calculated. The average rating was multiplied by 2.5 to obtain the final Dark Pattern Darkness Score. Doing this, the various types of Dark Patterns and their corresponding DPDSs could find themselves on a scale ranging from 0 to 10. Within this scale, a DPDS of 0 means a Dark Pattern is perceived to be not severe at all (or even 'harmless'), and a DPDS of 10 means a Dark Pattern has a very high severity impact.

Table 13: Dark Pattern types with their average rating and final DPDS.

Dark Pattern	Avg. Rating	DPDS
Hidden Costs	3.80	9.5
Roach Motel	3.87	9.5
Bait and Switch	3.77	9.4
Sneak into Basket	3.60	9.0
Disguised Ads	3.60	9.0
Nagging	3.53	8.8
Social Pyramid	3.48	8.7
Privacy Zuckering	3.45	8.6
Hidden Information	3.41	8.5
Trick Questions	3.38	8.5
Price Comparison Prevention	3.07	7.7
Forced Continuity	2.90	7.3
False Hierarchy	2.88	7.2
Intermediate Currency	2.80	7.0
Gamification	2.64	6.6
Preselection	2.33	5.8
Toying with Emotions	2.08	5.2

Table 13 provides the average ratings of the various Dark Pattern types, as well as their final Dark Pattern Darkness Score. Within the table, Dark Patterns are ordered by their DPDS, from highest (most severe) to lowest (least severe) score.

4.4.2 Grouping Dark Pattern Types

As the Dark Pattern Darkness Scores of the various Dark Pattern types were found to be relatively close to each other, defining harsh thresholds to categorise the Dark Patterns into various severity-impact categories (for example, 'high', 'moderate', 'low') seemed impossible. As such, the researcher decided to try and take another approach, by making use of *similarity*.

The goal of this approach is to cluster -or group- items (in this case: the Dark Pattern types) that have similar frequency distributions. One of the most common methods used for performing such an analysis with Likert scale data is *cosine similarity* (GeeksforGeeks, 2020; Stack Exchange, 2013). As such, a cosine similarity matrix was created using the DPDS questionnaire data. The matrix is provided in table 14. Within the matrix, 0 means no similarity, whereas 1 means that two Dark Patterns are 100% similar.

Table 14: Cosine Similarity Matrix.

	S01	**S02**	**S03**	**S04**	**S05**	**S06**	**S07**	**S08**	**S09**	**S10**	**S11**	**S12**	**S13**	**S14**	**S15**	**S16**	**S17**
S01	-	0.97	0.93	0.9	0.94	0.96	0.97	0.96	0.98	0.62	0.65	0.9	0.98	0.98	0.99	0.98	0.83
S02	0.97	-	0.83	0.85	0.88	1	1	1	0.97	0.49	0.54	0.79	0.99	0.97	0.98	0.98	0.77
S03	0.93	0.83	-	0.96	0.98	0.81	0.86	0.82	0.93	0.86	0.81	0.99	0.89	0.93	0.92	0.91	0.94
S04	0.9	0.85	0.96	-	0.98	0.85	0.89	0.85	0.94	0.87	0.85	0.96	0.9	0.94	0.93	0.93	0.99
S05	0.94	0.88	0.98	0.98	-	0.87	0.91	0.87	0.95	0.84	0.85	0.98	0.92	0.96	0.95	0.95	0.96
S06	0.96	1	0.81	0.85	0.87	-	1	1	0.97	0.48	0.52	0.78	0.99	0.97	0.98	0.98	0.77
S07	0.97	1	0.86	0.89	0.91	1	-	1	0.99	0.56	0.58	0.83	1	0.99	0.99	0.99	0.82
S08	0.96	1	0.82	0.85	0.87	1	1	-	0.97	0.49	0.53	0.78	0.99	0.97	0.98	0.98	0.77
S09	0.98	0.97	0.93	0.94	0.95	0.97	0.99	0.97	-	0.67	0.66	0.9	0.99	1	1	1	0.88
S10	0.62	0.49	0.86	0.87	0.84	0.48	0.56	0.49	0.67	-	0.92	0.89	0.59	0.68	0.65	0.64	0.93
S11	0.65	0.54	0.81	0.85	0.85	0.52	0.58	0.53	0.66	0.92	-	0.87	0.6	0.67	0.65	0.65	0.88
S12	0.9	0.79	0.99	0.96	0.98	0.78	0.83	0.78	0.9	0.89	0.87	-	0.85	0.9	0.89	0.89	0.95
S13	0.98	0.99	0.89	0.9	0.92	0.99	1	0.99	0.99	0.59	0.6	0.85	-	0.99	1	1	0.83
S14	0.98	0.97	0.93	0.94	0.96	0.97	0.99	0.97	1	0.68	0.67	0.9	0.99	-	1	1	0.89
S15	0.99	0.98	0.92	0.93	0.95	0.98	0.99	0.98	1	0.65	0.65	0.89	1	1	-	1	0.87
S16	0.98	0.98	0.91	0.93	0.95	0.98	0.99	0.98	1	0.64	0.65	0.89	1	1	1	-	0.87
S17	0.83	0.77	0.94	0.99	0.96	0.77	0.82	0.77	0.88	0.93	0.88	0.95	0.83	0.89	0.87	0.87	-

Within the above Similarity Matrix, 'Sxx' represents a situation related to a Dark Pattern type, with S01 representing 'Nagging', S02 'Roach Motel', S03 'Price Comparison Prevention', S04 'Intermediate Currency', S05 'Forced Continuity', S06 'Hidden Costs', S07 'Sneak into Basket', S08 'Bait & Switch', S09 'Hidden Information', S10 'Preselection', S11 'Toying with Emotions', S12 'False Hierarchy', S13 'Disguised Ads', S14 'Trick Questions', S15 'Social Pyramid', S16 'Privacy Zuckering', and S17 'Gamification'.

The cosine similarity matrix was used as input for hierarchical cluster analysis. Hierarchical clustering starts by treating each observation as a separate cluster. Then, it repeatedly executes the following two steps: (1) identify the two clusters that are closest together, and (2) merge the two most similar clusters. This iterative process continues until all the clusters are merged together. The main output of Hierarchical Clustering is a dendrogram, which shows the hierarchical relationship between the clusters (Bock, n.d.). The dendrogram of the current study is provided in figure 42.

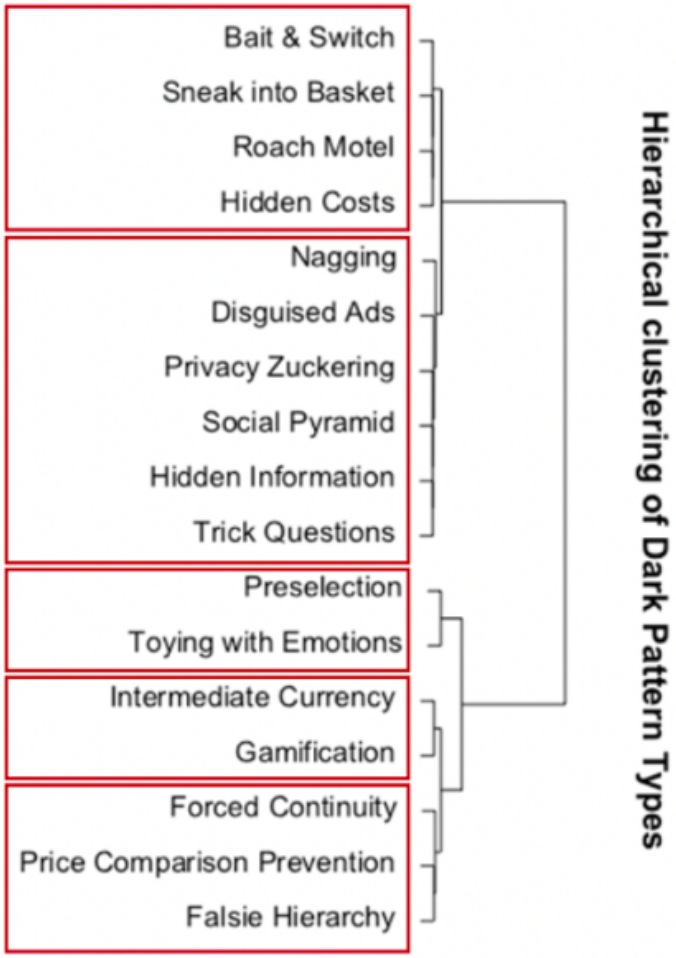


Figure 42: Dendrogram of Dark Pattern types. Dark Pattern types that the cosine matrix indicated to be similar to each other are grouped under the same 'branch' of the tree-like diagram. The red rectangles represent the final number of clusters.

Determining the optimal number of clusters is often a somewhat arbitrary choice (Jones, 2019). Within the current study, the researcher decided to go for a number of 5 clusters. By picking 5 clusters, each of the clusters is 'large enough' to include multiple Dark Patterns. By picking more clusters, this would not have been the case. Additionally, a number of 5 clusters enables other researchers (and users) to refer to the various Dark Pattern types on a relatively fine grained scale, whereas this would not have been possible with a number of less than 5 clusters.

4.4.3 Frequency of Encountering Dark Patterns

As already mentioned before, for each of the situations (or: Dark Pattern types) presented in the DPDS questionnaire, participants were also asked to indicate how frequently they had encountered the Pattern. Table 15 presents the mean of the responses participants provided on the 5-point likert scale ranging from (1) Never to (5) Very Often.

Table 15: Dark Pattern types with their average frequency rating.

Dark Pattern	**Mean**	**SD**
Preselection	4.11	1.01
Forced Continuity	3.85	1.05
Disguised Ads	3.64	1.16
Nagging	3.63	0.99
Hidden Costs	3.34	1.08
Intermediate Currency	3.00	1.56
Privacy Zuckering	2.86	1.25
Toying with Emotions	2.77	1.35
Gamification	2.71	1.39
Hidden Information	2.64	1.36
Price Comparison Prevention	2.54	1.23
Roach Motel	2.36	1.11
Trick Questions	2.28	1.10
Social Pyramid	2.27	1.21
Bait and Switch	2.15	1.19
False Hierarchy	2.07	1.07
Sneak into Basket	1.64	0.88

No correlation was found between 'frequency of encountering a Dark Pattern' and the 'severity rating of a Dark Pattern': $r(-0.28) = .34$, $p = 0.28$. In other words, Dark Patterns that received high severity ratings were not necessarily the ones that were most encountered -or the other way around.

In figure 43, a scatter plot is presented, showing the non-existent correlation between 'Frequency of Encounter' and 'Severity Rating'.

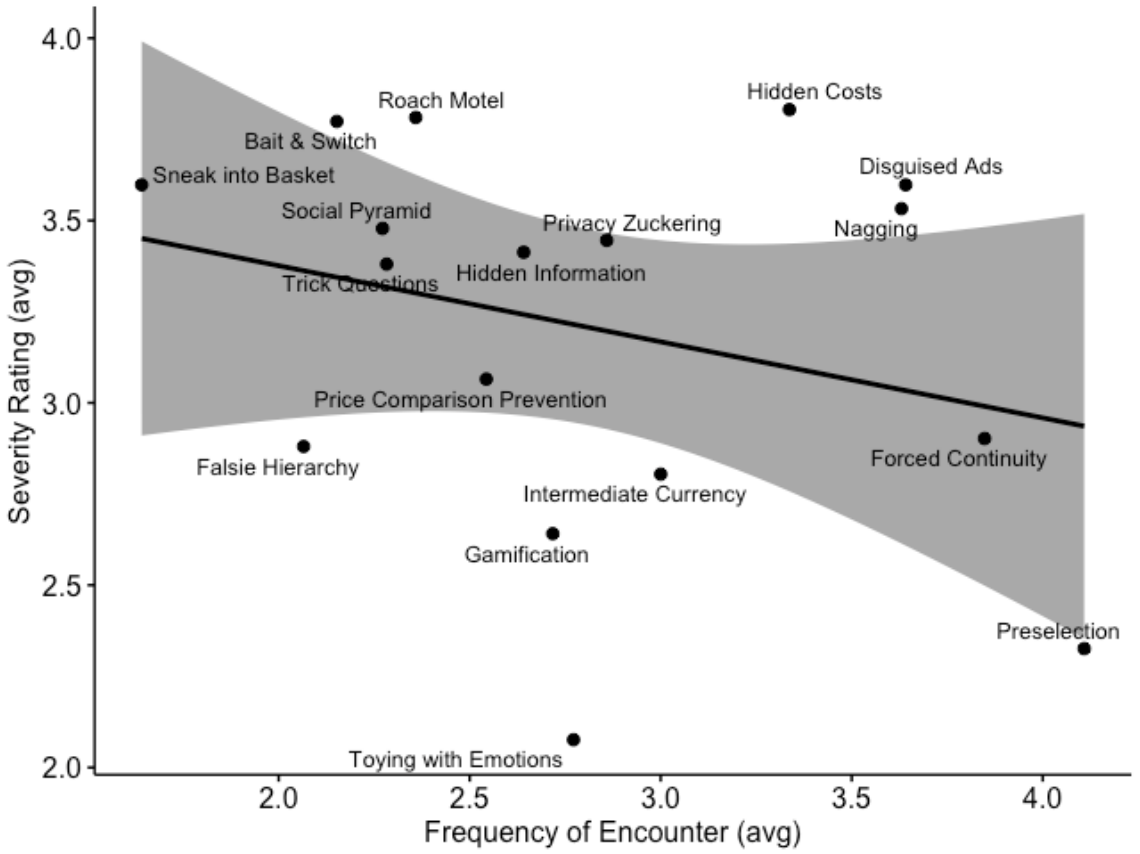


Figure 43: Scatter plot representing each of the Dark Pattern types in terms of their average Severity Rating and average Frequency of Encounter.

4.4.4 Pre-knowledge: Different Ratings for Version A and B?

Although Study 3 treated all participants as one group, it could have been the case that participants had different pre-knowledge -or were differently primed- based on whether they performed in either the 'Dark' or 'Bright' condition of Studies 1 and 2. As such, this subsection will investigate whether 'version' had an effect on the Dark Pattern Darkness Scores provided by participants.

Table 16 presents the average rating and final DPDS for each of the Dark Pattern types, for both Version A and B. The table shows that average ratings (and therefore DPDSs) are fairly similar for both versions. However, to be absolutely sure 'version' did not have an effect on the final DPDSs (as presented in table 13), some statistical tests were performed as well.

Table 16: Average Severity Ratings and final DPDSs for each of the Dark Pattern types, for both Version A and B of Studies 1 and 2. Whereas Version A represents the *no* Dark Patterns condition, Version B represents the condition in which Dark Patterns *were* presented to participants in the previous studies. The last column of the table shows the results of the Mann Whitney U-test.

	Version A		**Version B**		**Mann Whitney Test**		
Dark Pattern	**Avg. Rating**	**DPDS**	**Avg. Rating**	**DPDS**	**U**	**z**	**p**
Hidden Costs	3.9	9.7	3.7	9.3	980.0	-.950	.342
Roach Motel	3.8	9.4	3.8	9.5	1035.0	-.267	.790
Bait & Switch	3.8	9.5	3.7	9.3	1013.0	-.533	.594
Sneak into Basket	3.5	8.8	3.7	9.2	1029.5	-.289	.772
Disguised Ads	3.6	9.1	3.6	8.9	963.0	-.928	.353
Nagging	3.6	9.0	3.5	8.7	1012.0	-.423	.672
Social Pyramid	3.6	9.1	3.3	8.3	874.5	-1.684	.092
Privacy Zuckering	3.4	8.5	3.5	8.7	1034.5	-.215	.829
Hidden Information	3.6	8.9	3.3	8.2	856.5	-1.817	.070
Trick Questions	3.7	9.1	3.1	7.8	719.5	-3.028	.002*
Price Comparison Prevention	3.1	7.8	3.0	7.6	1013.0	-.372	.710
Forced Continuity	2.8	7.0	3.0	7.6	937.0	-.996	.319
False Hierarchy	3.0	7.6	2.7	6.8	944.5	-.928	.353
Intermediate Currency	2.9	7.3	2.7	6.7	917.0	-1.154	.248
Gamification	2.7	6.7	2.6	6.5	1007.5	-.409	.683
Preselection	2.5	6.4	2.1	5.3	816.0	-1.950	.051
Toying with Emotions	2.4	6.1	1.7	4.3	743.0	-2.516	.012*

As none of the data was normally distributed, two-tailed Mann Whitney U-tests were used to see whether participants that performed in different versions scored the various Dark Pattern types differently in terms of severity. A significant difference was found for two of the Dark Pattern types:

- For the *'Toying with Emotions'* Dark Pattern, Severity scores (DPDSs) of participants participating in Version A of Studies 1 and 2 (no Dark Patterns in UI) were higher than those of participants in Version B of Studies 1 and 2 (Dark Patterns in UI). The Mann Whitney U-test indicated that this difference was statistically significant: U ($N_{VersionA}$ = 46, $N_{VersionB}$ = 46) = 743, z = -2.516, p = 0.012.
- For the *'Trick Questions'* Dark Pattern, Severity scores (DPDSs) of participants participating in Version A of Studies 1 and 2 (no Dark Patterns in UI) were higher than those of participants in Version B of Studies 1 and 2 (Dark Patterns in UI). The Mann Whitney U-test indicated that this difference was statistically significant: U ($N_{VersionA}$ = 46, $N_{VersionB}$ = 46) = 719.5, z = -3.028, p = 0.002.

As for *only* 2 out of the 17 Dark Pattern types a significant difference in severity scores was found between Version A and B, we can assume that the version in which participants participated in Study part 1 and 2 has not really played a role.

5 Conclusion

This thesis looked at Dark Patterns from a user's perspective. As the research possibilities within this perspective were endless, the researcher could not help but conduct multiple studies. As such, the current study was threefold. It investigated a user's view upon the malicious Dark Pattern strategies -in terms of their level of severity- and showed how user's feel towards interfaces where Dark Pattern strategies are present. Additionally, a user's take upon the darkness (or brightness) of the UI design of a website as a whole was captured. Several conclusions can be drawn from this study.

5.1 Study 1: Performing an Experiment on User Experience

The results of the User Experience experiment have shown that both the 'Level of Satisfaction' and 'Recommendation Score', the two measurable variables related to User Experience, were influenced by Dark Patterns -at least to some extend.

Participants who *did not* encounter Dark Patterns during their shopping task (Version A) provided the e-commerce website with significantly higher Recommendation Scores than participants who *did* encounter Dark Patterns during their shopping task (Version B). As such, we can conclude that overall, Version A and B of the experiment lead to different user experiences, with the experience being worse in Version B: When Dark Pattern are implemented in the interface.

However, results of the seven UX pop-ups showed that only two of the web pages within the shopping task really contributed to this difference in User Experience. Compared to the Dark Patterns web shop (Version B), participants in the *no* Dark Patterns web shop (Version A) provided higher Level of Satisfaction ratings for the 'shopping cart' page and 'place order' page. For all the other pages, Level of Satisfaction rates were fairly similar.

The Dark Patterns implemented within the 'shopping cart' page were: (1) Sneak into Basket, and (2) Toying with Emotions (countdown timer). Dark Patterns implemented within the 'place order' page were: (1) Hidden Costs, (2) Hidden Information, (3) Preselection, and (4) Trick Questions.

Looking at the Textual Explanations upon the Recommendation Score -which served as additional insights- we saw that a high percentage of participants performing in Version B of the experiment mentioned a series of Dark Patterns to have contributed to a (fairly) low Recommendation score. These were the exact same Dark Patterns that were found to be contributing to a significant difference in User Experience (e.g., the Dark Patterns included in the 'shopping cart' and 'place order' pages).

A potential reason why a significant difference was found only for two of the task-related web pages, can be found within research on *Dark Pattern Blindness* (section 2.7.3). It might be the case that participants *did not notice* the Dark Pattern types presented on the other five web pages, and therefore did not indicate their 'level of satisfaction' for these pages to be diminished. This explanation can be substantiated by the fact that the Dark Pattern types present on the non-significant pages were not mentioned in the Textual Explanations upon the Recommendation Score as well.

5.2 Study 2: Creating a 'System Darkness Scale'

The result of the System Darkness Questionnaire is the System Darkness Scale (SDS). The final SDS consists of 5 questionnaire items, all related to some aspect of darkness.

The five items that were selected out of the pool of 25 potential items were found to all be measuring the same construct ('darkness'), providing the SDS with a good level of internal consistency. The items were also all capable of discriminating between the responses provided for a 'bright' and 'dark' version of a system, thereby providing the SDS with the competence of creating an accurate representation of the 'darkness' (SDS scores towards 100) or 'brightness' (SDS scores towards 0) of a system.

5.3 Study 3: Assigning a Severity Score to Dark Patterns

The result of the Dark Pattern Darkness Score Questionnaire (Study 3) provided each of the 17 Dark Patterns -as classified by Gray et al. (2018)- with a severity score.

When ordering the Dark Pattern types based on their DPDS, we could see that participants perceived the following Dark Patterns to be the 'most severe' (DPDS of at least '9'): (1) Hidden Costs, (2) Roach Motel, (3) Bait & Switch, (4) Sneak into Basket. What is interesting to note here, is that at least two of these Patterns (Hidden Costs & Sneak into Basket) have a direct relation to *monetary aspects*.

Participants perceived (1) Toying with Emotions, and (2) Preselection to be the 'least severe' Dark Pattern types (DPDS lower than '6'). Although we are unsure why Toying with Emotions is not perceived to be severe, for Preselection there is a relatively clear explanation. As we saw in figure 43, Preselection is one of the Dark Patterns that is mostly encountered by users. The frequent encounter with Preselection might have led to users 'getting used' to it, resulting in less frustration with the pattern on the long term.

5.4 (Dis)similarities of Studies 1, 2 & 3

5.4.1 Relating Study 1 & Study 2

None of the potential SDS questionnaire items in Study 2 was formulated around a specific Dark Pattern type. Rather, they were formulated in a generalizable manner, in order for future SDS users to be able to capture the Darkness of a variety of systems, having a variety of sub-sets of Dark Pattern types.

Although this was the case, for most of the SDS items (at least) one of the Dark Pattern types can be related to it. For example, one of the items selected for the final SDS is about 'spending more money than anticipated'. A Dark Pattern that can be related to this item is 'Hidden Costs': If a user makes it to the end of a certain (shopping) process and finds out more costs are involved than was previously shown, he will most likely still finish the process, as much effort is put in the process already. As a result, the user is spending more money than originally anticipated.

But one can think of 'Toying with Emotions' here as well: If a user is presented with a text stating that a certain product is in 'high demand', he might decide to buy *this* product instead of a cheaper version, as he thinks products in high demand are more popular, and therefore better. When buying the 'more expensive' product, the user is -again- spending more money than anticipated.

When comparing the results of Study 1 and Study 2, we are able to relate some of the Dark Patterns that contributed to a negatively influenced User Experience (for the 'shopping cart' and 'place order' page) to some of the final SDS questionnaire items. In the previous paragraph, we already explained how the 'Hidden Costs' Dark Pattern can be related to the 'The system pushed me into spending more money than I originally anticipated' item. The 'Sneak into Basket' Dark Pattern, in turn, can be associated with the 'The system performed certain actions I was not aware of' item.

Figure 44 visualises the potential relationship between the results of Study 1 and 2.

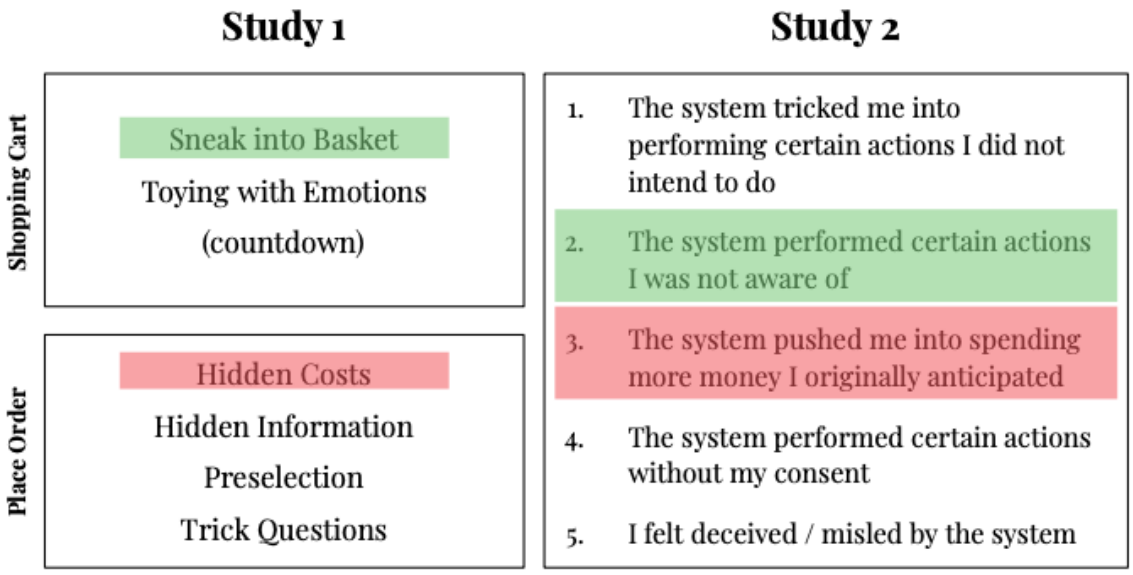


Figure 44: Relationship between the results of Study 1 and 2.

The fact that some of the Dark Patterns that were found to be contributing to a 'significant' difference in Study 1 are (to some extend) represented in the final SDS of Study 2 might come as no surprise, as the e-commerce websites of Study 1 were used as input for Study 2. We will discuss this in more detail in the Discussion section.

5.4.2 Relating Study 1 & Study 3

For relating Study 1 and 3, it might be interesting to look at the Dark Pattern types that were found to be significantly influencing User Experience in Study 1, and see what Dark Pattern Darkness Score (DPDS) participants provided them with in Study 3:

- Hidden Costs: 9.5
- Sneak into Basket: 9.0
- Hidden Information: 8.5
- Trick Questions: 8.5
- Preselection: 5.8
- Toying with Emotions: 5.2

Figure 45 visualises the potential relationship between the results of Study 1 and 3 in another way.

What is remarkable, is that out of the six Dark Patterns that were included in the significant 'shopping cart' and 'place order' pages, two of them received the lowest DPDSs: 'Preselection' and 'Toying with emotions'. There might be multiple explanations for this.

The most obvious explanation would be that 'Preselection' and 'Toying with Emotions' did not really contribute to the significant result found in Study 1 at all. As multiple Dark Patterns were present in both 'significant pages', it is not unlikely that the four Dark Patterns other than 'Preselection' and 'Toying with Emotions' have had a bigger stake in the final significant result for User Experience.

This possibility is reinforced by the fact that the DPDSs for the four other Dark Patterns are (relatively) high. 'Hidden Costs' and 'Sneak into Basket' both belonged to the 4 Dark Patterns with the highest severity scores. 'Hidden Information' and 'Trick Questions' received a somewhat lower DPDS (finding themselves on the 9th and 10th place in the DPDS ranking), but still were perceived to be way more severe than 'Preselection' and 'Toying with Emotions'.

At least for the 'Preselection' Dark Pattern, the scenario described above seems not unlikely. As table 15 showed, out of all Dark Patterns, participants indicated to encounter 'Preselection' the most frequently. As such, it might be the case that users are already *so* familiar with this Dark Pattern, that they don't even find this pattern annoying or severe anymore. Hence the low DPDS, and hence the possibility that 'Preselection' did not really contribute to the significant difference in User Experience.

For the 'Toying with Emotions' Dark pattern, the scenario described above is a bit more uncertain. Reason for this is that this particular pattern has different instances. Two different instances were used for Study 1 and Study 3. For Study 1, 'Toying with Emotions' took the form of a 'countdown timer'. The situation in Study 3 described the use of 'emotional wording' in offer pop-ups. Whereas users indicated 'emotional wording' to have low severity impact, it might be possible that they would have rated other instances of 'Toying with Emotions', such as the 'countdown timer', differently. We will discuss this topic in more detail in the Discussion section.

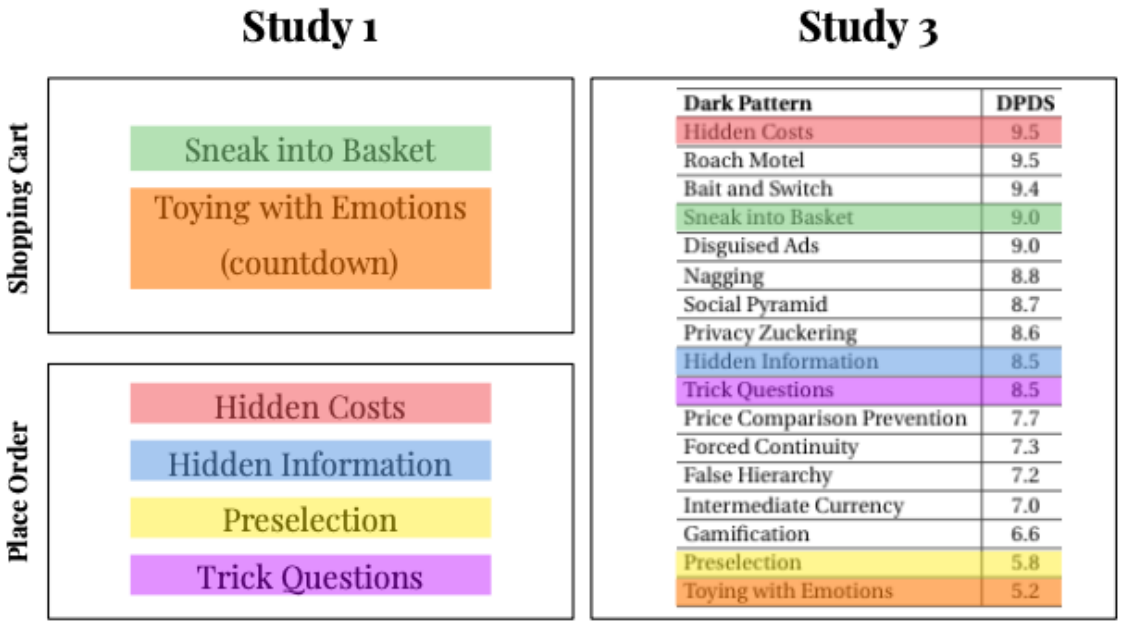


Figure 45: Relationship between the results of Study 1 and 3.

The reason why the 'Roach Motel' and 'Bait & Switch' Dark Patterns are not included in the Dark Patterns that have lead to a significant difference in User Experience in Study 1 (whereas they are perceived to be very severe in Study 3!) is simply because they were not included in the e-commerce web shop -as they did not really match this platform.

5.4.3 Relating Study 2 & Study 3

It was already mentioned before that most Dark Pattern types can be associated with at least one of the potential SDS questionnaire statements. When looking at the final set of SDS statements, and comparing this set with the four Dark Pattern types that participants perceived to be the most 'severe', we can see that 'Hidden Costs' is represented by the 'The system pushed me into spending more money than I originally anticipated' statement and 'Sneak into Basket' is represented by the 'The system performed certain actions I was not aware of' statement.

As such, one can conclude that these SDS items might be good representatives of Darkness, as they capture some of the 'darkest' patterns. Whenever these Dark Patterns are included in the user interface of a system, participants will probably have high agreement to these items, boosting the SDS score, resulting in a score higher up the Darkness scale.

Figure 46 visualises the potential relationship between the results of Study 2 and 3.

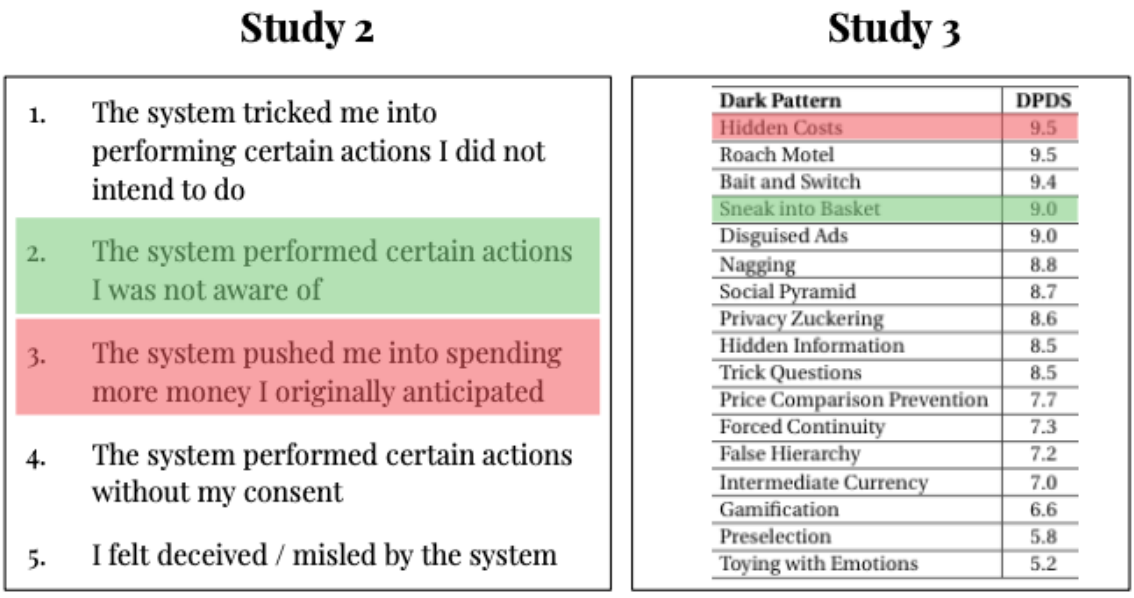


Figure 46: Relationship between the results of Study 2 and 3.

6 Discussion

This section discusses the opportunities and limitations of the current study. It also provides researchers within the field of HCI with some insights upon possibilities for future work.

6.1 Opportunities

Since little research has been performed on the *perception of users* upon Dark Patterns, the current study forms a valuable contribution to the overall body of knowledge for the domain of Human Computer Interaction, or more specifically: UI design.

While previous studies found Dark Patterns to be very effective in misleading users -thereby creating positive assets for the online companies at hand-, the current study showed that Dark Patterns lead to *negative effects* for companies as well. Negative user satisfaction, or a negative user experience, might for example result in users not returning to the platform.

As such, this research hopefully shows companies the other ('dark') side of the medal of using Dark Patterns in the UI of their platforms. In the most optimal scenario, this insight leads companies to removing malicious strategies from their platforms, or prevents them from even starting to use these strategies in the first place.

By creating more awareness upon the different negative effects of Dark Patterns among both (online) platform users and suppliers, we will hopefully return to an online world in which the user is put at the centre of attention again, and in which business benefits are *earned* by treating users with care.

6.2 Limitations

Although the current study provides lots of new insights concerning Dark Patterns, it has some limitations that are worth mentioning as well. In what follows, the limitations for each of the three study parts will be discussed.

6.2.1 Study 1: Performing an Experiment on User Experience

The first experiment measured whether the presence of Dark Patterns has an effect on User Experience. The results of the referral pop-up showed that -overall- the use of malicious strategies leads to a decrease in User Experience. However, the current research only investigated this effect for one specific system: E-commerce web shops. As such, it is not clear whether the found result holds for -or is generalizable over other- (online) systems, such as gaming applications or ticket booths.

A second limitation of the User Experience experiment is related to the number of Dark Patterns that was included in the various task-related web pages of the e-commerce web shop. In an attempt to include as many Dark Patterns as possible, some web pages contained multiple Dark Patterns, whereas others contained only one Dark Pattern (or even none).

Both pages that were found to lead to a significantly different User Experience contained multiple Dark Patterns. We can state with certainty that the *set* of Dark Patterns present on each of these pages added to this difference. However, what we do not know is what the specific contribution of each of the separate Dark Pattern types to this result is. In order to determine the effect of each of the 17 different Dark Pattern types on User Experience, Dark Pattern types should probably not be presented to the participant at the same time. This might be an interesting research topic for future work.

A final limitation of Study 1 has to do with the way some data was gathered. For gathering 'Level of Satisfaction' ratings of participants, a UX pop-up was presented to them right after browsing each of the task-related pages. This form of gathering data was chosen upon, as the ratings of participants were thought to be most accurate when they provided them right after interacting with each of the pages. However, the downside of providing participants with various experimental pop-ups is that it might have impaired the (shopping) task flow.

A post-test (containing screenshots of the web shop for each page) might be another possibility to gather Satisfaction data. However, in this case, user experience ratings of participants might be less accurate, as participants need to *remember* how they felt at each interaction step.

6.2.2 Study 2: Creating a 'System Darkness Scale'

The second study took the first steps towards creating a System Darkness Scale: A Likert scale questionnaire which can be used to evaluate the 'darkness' of a certain system. One of the biggest limitations of this Study, again, has to do with the *specific* system being used to develop the SDS: the e-commerce web shop of Study 1. Results of the final SDS (e.g., the five selected questionnaire items) are based solely on participants interaction with this e-commerce website.

Within the B version of this website, not all existing types of Dark Patterns were present: Only e-commerce related Dark Patterns were implemented. As such, if you would, for example, ask participants to fill out the potential pool of 25 SDS statements after interacting with a 'Bright' and 'Dark' version of a gaming application, you might end up with totally different results -and therefore with a totally different final SDS questionnaire.

Another limitation is the fact that the A and B version of the e-commerce system were classified as 'bright' and 'dark' without running checks. As such, although the systems were said to be at the extreme ends of the spectrum of attitude that needed to be captured, this was not specifically verified. A better, future approach would be to run a pre-study for selecting the study material (e.g., opposite systems).

A third limitation of Study 2 is related to assembling the pool of potential SDS questionnaire items. Due to the limited time available for performing the current study, the researcher had to come up with the list of potential items herself. This could have resulted in an *incomplete* pool of statements. A possibly more elegant way of assembling a list of statements would have been to perform some brainstorm sessions with experts in the field.

A final limitation of Study 2 is more related to the content of the final SDS questionnaire, rather than the set-up of the research leading to its creation. A less satisfactory feature of the SDS is that it (accidentally) became unbalanced: The items on the final scale are all worded in the same direction (e.g., the 'dark' direction). As such, the SDS becomes prone to *response bias*. This bias can be caused by the fact that respondents (e.g., users of the scale) do not have to think about each statement.

More preferably, the SDS would be a balanced scale, in which at least some (and preferably half) of the items would be worded in the 'bright' direction. By alternating positive and negative items, respondents have to read each statement, and make an effort to think whether they agree or disagree with it.

As already mentioned before, the SDS turned out to be unbalanced *accidentally*. Within the pool of 25 potential statements, statements were formulated in both directions. However, the 5 selected statements turned all out to be 'dark' ones. As the formulation of statements on a Likert Scale can have a major effect on the responses of respondents, the researcher decided not to change half of the selected items into its opposite equivalent.

6.2.3 Study 3: Assigning a Severity Score to Dark Patterns

The third study assigned each of the 17 Dark Pattern types a certain severity (or 'darkness') score. The biggest limitation of this study probably is the limited number of Dark Pattern situations that was presented to the participants. More specifically, for each of the 17 Dark Pattern types, only one example situation was formulated.

However, some of the Dark Pattern types have more than one instance. 'Toying with Emotions', for example, can take the form of 'Emotional Wording, a 'Countdown Timer', or a 'Limited Stock Message'. It might be the case that participants would provide significantly *different* severity scores (DPDS) for the *different* instances of the *same* Dark Pattern. As such, it might be questionable whether we can capture the severity of a Dark Pattern with multiple instances with a single score.

Another limitation of Study 3 is that Dark Patterns were described within a rather *specific* situation. For example, the 'nagging' Dark Pattern was described within a Social Media setting. Participants had to imagine themselves encountering the Dark Pattern on the Social Media platform 'Instagram'. Some participants might have find it hard to replace themselves in this situation (for example, because they have no Instagram account). As such, their DPDS scores might have been inaccurate or not representable.

It might be a good idea to create -or formulate- various different situations for the same Dark Pattern type (or even for their multiple instances), in order for participants to relate themselves to at least one of the situations, and get a more accurate severity score (DPDS).

6.3 Future Research

As still little research has been performed within the domain of Dark Patterns, lots of research possibilities remain present. For example, further research could consider using eye trackers in Dark Pattern experiments. Eye tracking data may provide researchers with further insights into how a participants perceives a certain (maliciously crafted) system. Questions that may be answered by using eye trackers, for example, can be "In systems containing Dark Patterns, did the participant even *notice* the Dark Pattern?", "If a Dark Pattern was noticed, how much time did the participant spent looking at the Pattern?", and "Which Dark Pattern types received the most attention? Are these the Dark Patterns that also received the highest severity scores (DPDS)?".

Another research possibility, which is more directly related to the current study, is to validate and build upon the final System Darkness Scale (SDS) that was created in Study part 2. By using various different systems, various different combinations of Dark Pattern types, and probably an even bigger pool of statements to choose from, the SDS has the potential to develop into an even better tool to evaluate the darkness of systems. In other words, further research is desired to turn the SDS into a tool that becomes widely accepted within the HCI community.

Future research could also investigate whether it is possible to relate the Dark Pattern Darkness Scores (DPDS) -provided for each of the Dark Pattern types- with the score resulting from the System Darkness Scale (SDS). Within the current study, we saw that each Dark Pattern type is perceived differently in terms of its severity. Different combinations of Dark Pattern types will therefore probably lead to different 'overall' severity (or darkness) scores, which is exactly what the SDS measures. It would be interesting to see if this potential relationship can be captured by some kind of formula.

Another possibility to relate the SDS to the DPDS is to use the DPDS severity scores that were assigned to the various Dark Pattern types as a means to create a weighted SDS, where the responses to some items have a higher weighting in -or do contribute more to- the final SDS 'darkness' score than others. For example, one might investigate whether the response to SDS item 3 -'The system pushed me into spending more money than I originally anticipated'- should have a higher weighting than the other SDS items, as we saw that some of the 'more severe' Dark Patterns are related to monetary aspects.

Finally, future research should take into account all the limitations of the current study, which are provided in section 6.2.

References

Acquisti, A., Adjerid, I., Balebako, R., Brandimarte, L., Cranor, L. F., Komanduri, S., Leon, P. G., Sadeh, N., Schaub, F., Sleeper, M. et al. (2017). Nudges for privacy and security: Understanding and assisting users' choices online. *ACM Computing Surveys (CSUR)*, *50*(3), 1–41.

Arteaga, S. M., Kudeki, M. & Woodworth, A. (2009). Combating obesity trends in teenagers through persuasive mobile technology. *ACM SIGACCESS Accessibility and Computing*, (94), 17–25.

Bangor, A., Kortum, P. T. & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction*, *24*(6), 574–594.

Batterbee, I. (2020). *Don norman's seven important questions of user interaction*. Retrieved December 15, 2020, from https://uxdesign.cc/ux-psychology-principles-seven-important-questions-960579272880

Berezhnoi, R. (2019). *What is ui design and why is it important?* Retrieved December 23, 2020, from https://f5-studio.com/articles/what-is-user-interface-design-and-why-is-it-important/

Bock, T. (n.d.). *What is hierarchical clustering?* Retrieved June 8, 2021, from https://www.displayr.com/what-is-hierarchical-clustering/

Bösch, C., Erb, B., Kargl, F., Kopp, H. & Pfattheicher, S. (2016). Tales from the dark side: Privacy dark strategies and privacy dark patterns. *Proceedings on Privacy Enhancing Technologies*, *2016*(4), 237–254.

Brignull, H. (2011). *Dark patterns: Deception vs. honesty in ui design*. Retrieved January 20, 2021, from https://alistapart.com/article/dark-patterns-deception-vs.-honesty-in-ui-design

Brignull, H., Miquel, M., Rosenberg, J. & Offer, J. (2015). Dark patterns-user interfaces designed to trick people.

Brooke, J. (1996). Sus: A "quick and dirty'usability. *Usability evaluation in industry, 189*.

Brownlee, J. (2016). *Why dark patterns won't go away*. Retrieved January 20, 2021, from https://www.fastcompany.com/3060553/why-dark-patterns-wont-go-away

Buley, L. (2013). *The user experience team of one: A research and design survival guide*. Rosenfeld Media.

Canfield, D. d. S. & Basso, K. (2017). Integrating satisfaction and cultural background in the customer journey: A method development and test. *Journal of International Consumer Marketing*, *29*(2), 104–117.

Chapin, B. (n.d.). *Customer journey maps – what they are and how to build one*. Retrieved January 27, 2021, from https://www.toptal.com/designers/product-design/customer-journey-maps

Chiu, M.-C., Chang, S.-P., Chang, Y.-C., Chu, H.-H., Chen, C. C.-H., Hsiao, F.-H. & Ko, J.-C. (2009). Playful bottle: A mobile social persuasion system to motivate healthy water intake. *Proceedings of the 11th international conference on Ubiquitous computing*, 185–194.

Clark, T. (2008). We're over-researched here!' exploring accounts of research fatigue within qualitative research engagements. *Sociology*, *42*(5), 953–970.

Consolvo, S., Everitt, K., Smith, I. & Landay, J. A. (2006). Design requirements for technologies that encourage physical activity. *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 457–466.

Consolvo, S., Klasnja, P., McDonald, D. W. & Landay, J. A. (2009). Goal-setting considerations for persuasive technologies that encourage physical activity. *Proceedings of the 4th international Conference on Persuasive Technology*, 1–8.

Convertize. (n.d.). *Scarcity effect*. Retrieved January 20, 2021, from https://www.convertize.com/glossary/scarcity-effect/

De Rosis, F., Carofiglio, V., Grassano, G. & Castelfranchi, C. (2003). Can computers deliberately deceive? a simulation tool and its application to turing's imitation game. *Computational Intelligence*, *19*(3), 235–263.

Dennis, A. R., Yuan, L., Feng, X., Webb, E. & Hsieh, C. J. (2020). Digital nudging: Numeric and semantic priming in e-commerce. *Journal of Management Information Systems*, *37*(1), 39–65.

Di Geronimo, L., Braz, L., Fregnan, E., Palomba, F. & Bacchelli, A. (2020). Ui dark patterns and where to find them: A study on mobile applications and user perception. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14.

Edraw. (n.d.). *Online shopping customer journey map template*. Retrieved January 21, 2021, from https://www.edrawsoft.com/template-online-shopping-customer-journey-map.html?utm_source=youtube&utm_medium=social&utm_campaign=youtube_video&utm_term=customer-jouney-2019&utm_content=link_ed_max_en_19063345_2021-01-19

Field, A. (2013). *Discovering statistics using ibm spss statistics*. sage.

Fogg, B. (2003). Introduction: Persuasion in the digital age. *Persuasive Technology: Using Computers to Change What We Think and Do*, 1–13.

Forbruker Radet. (2021). *Amazon manipulates customers to stay subscribed*. Retrieved January 28, 2021, from https://www.forbrukerradet.no/news-in-english/amazon-manipulates-customers-to-stay-subscribed/

Forbrukerrådet. (2018). Deceived by design, how tech companies use dark patterns to discourage us from exercising our rights to privacy. *Norwegian Consumer Council Report*.

GeeksforGeeks. (2020). *Item-to-item based collaborative filtering*. Retrieved June 8, 2021, from https://www.geeksforgeeks.org/item-to-item-based-collaborative-filtering/

Gilovich, T., Griffin, D. & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge university press.

Gray, C. M., Kou, Y., Battles, B., Hoggatt, J. & Toombs, A. L. (2018). The dark (patterns) side of ux design. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14.

Greenberg, S., Boring, S., Vermeulen, J. & Dostal, J. (2014). Dark patterns in proxemic interactions: A critical perspective. *Proceedings of the 2014 conference on Designing interactive systems*, 523–532.

Gubaidulin, I. (2016). *Designing for persuasion, emotion and trust*. Retrieved December 11, 2020, from https://uxdesign.cc/designing-for-persuasion-emotion-and-trust-cdac44c61d53

Gunnarsson, E. (2020). Log in or sign up.

Howard, T. (2014). Journey mapping: A brief overview. *Communication Design Quarterly Review*, *2*(3), 10–13.

Interaction Design Foundation. (n.d.). *Customer journey maps*. Retrieved January 27, 2021, from https://www.interaction-design.org/literature/topics/customer-journey-map

Interaction Design Foundation. (2002). *Human-computer interaction (hci)*. Retrieved December 23, 2020, from https://www.interaction-design.org/literature/topics/human-computer-interaction

Interaction Design Foundation. (2020a). *Usability*. Retrieved December 28, 2020, from https://www.interaction-design.org/literature/topics/usability

Interaction Design Foundation. (2020b). *User experience (ux) design*. Retrieved December 28, 2020, from https://www.interaction-design.org/literature/topics/ux-design

International Organization for Standardization. (2019). *Iso 9241-210, ergonomics of human-system interaction - part 210: Human-centred design for interactive systems*. Retrieved December 28, 2020, from https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-2:v1:en

Johnson, E. J., Bellman, S. & Lohse, G. L. (2002). Defaults, framing and privacy: Why opting in-opting out. *Marketing Letters*, *13*(1), 5–15.

Jones, T. W. (2019). *Document clustering*. Retrieved June 8, 2021, from https://cran.r-project.org/web/packages/textmineR/vignettes/b_document_clustering.html

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Karr, A. (2014). *Ethical design* (tech. rep.). Technical Report. Interactions ACM, https://interactions. acm. org/blog/view . . .

Keith, J. (2017). *Hooked and booked*. Retrieved January 20, 2021, from https://adactio.com/journal/%2013109

Kim, G. (2015). *Human-computer interaction*. Auerbach Publications.

Kitsing, R. (2018). Dark patterns of web design.

Láng, L. & Pudane, P. D. (2019). Deceptive interfaces: A case study on amazon's account deletion navigation and its effects on user experience.

Lin, J. J., Mamykina, L., Lindtner, S., Delajoux, G. & Strub, H. B. (2006). Fish'n'steps: Encouraging physical activity with an interactive computer game. *International conference on ubiquitous computing*, 261–278.

Luguri, J. & Strahilevitz, L. J. (2021). Shining a light on dark patterns. *Journal of Legal Analysis*, *13*(1), 43–109.

Maier, M. (2019). Dark patterns–an end user perspective.

Marquez, J. J., Downey, A. & Clement, R. (2015). Walking a mile in the user's shoes: Customer journey mapping as a method to understanding the user experience. *Internet Reference Services Quarterly*, *20*(3-4), 135–150.

Mathur, A., Acar, G., Friedman, M. J., Lucherini, E., Mayer, J., Chetty, M. & Narayanan, A. (2019). Dark patterns at scale: Findings from a crawl of 11k shopping websites. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–32.

McKay, E. N. (2013). *Ui is communication: How to design intuitive, user centered interfaces by focusing on effective communication*. Newnes.

Medium.com. (2017). *Chinese shoe company tricks people into swiping instagram ad with fake strand of hair*. Retrieved January 31, 2021, from https://medium.com/shanghaiist/chinese-shoe-company-tricks-people-into-swiping-instagram-ad-with-fake-strand-of-hair-54d8a2d8ec1d

Meske, C. & Potthoff, T. (2017). The dinu-model–a process model for the design of nudges.

Micheaux, A. & Bosio, B. (2019). Customer journey mapping as a new way to teach data-driven marketing as a service. *Journal of Marketing Education*, *41*(2), 127–140.

Mittone, L. & Savadori, L. (2009). The scarcity bias. *Applied Psychology*, *58*(3), 453–468.

Nagda, Y. (2020). Analyzing ocular parameters to investigate effect of dark patterns on hci.

Nevala, E. (2020). Dark patterns and their use in e-commerce.

Nodder, C. (2013). *Evil by design: Interaction design to lead us into temptation*. John Wiley & Sons.

Norman, D. (2013). *The design of everyday things: Revised and expanded edition*. Basic books.

Nouwens, M., Liccardi, I., Veale, M., Karger, D. & Kagal, L. (2020). Dark patterns after the gdpr: Scraping consent pop-ups and demonstrating their influence. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13.

Paay, J., Kjeldskov, J., Skov, M. B., Srikandarajah, N. & Brinthaparan, U. (2015). Quittylink: Using smartphones for personal counseling to help people quit smoking. *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 98–104.

Peng, W. (2009). Design and evaluation of a computer game to promote a healthy diet for young adults. *Health communication*, *24*(2), 115–127.

Preece, J., Sharp, H. & Rogers, Y. (2015). *Interaction design: Beyond human-computer interaction*. John Wiley & Sons.

Rosenbaum, M. S., Otalora, M. L. & Ramırez, G. C. (2017). How to create a realistic customer journey map. *Business Horizons*, *60*(1), 143–150.

Sauro, J. (2018). *The user experience of flower websites*. Retrieved January 3, 2021, from https://measuringu.com/flower-ux/

Schaffer, E. (2009). *Beyond usability: Designing web sites for persuasion, emotion, and trust*. Retrieved December 11, 2020, from https://www.uxmatters.com/mt/archives/2009/01/beyond-usability-designing-web-sites-for-persuasion-emotion-and-trust.php

Shopify. (2019). *Dark patterns: 12 tricks you should never use in your products*. Retrieved January 28, 2021, from https://www.shopify.com/partners/blog/dark-patterns

Sommerer, C. & Mignonneau, L. (2008). *The art and science of interface and interaction design (vol. 1)* (Vol. 141). Springer.

Stack Exchange. (2013). *Cluster analysis of ordinal variables (likert scale)*. Retrieved June 8, 2021, from https://stats.stackexchange.com/questions/56479/cluster-analysis-of-ordinal-variables-likert-scale

Sunstein, C. R. (2018).
better off, as judged by themselves": A comment on evaluating nudges. *International Review of Economics*, *65*(1), 1–8.

Thaler, R. H. & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

Trochim, W. M. (2021). *Likert scaling*. Retrieved June 17, 2021, from https://conjointly.com/kb/likert-scaling/

Tversky, A. & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *science*, *211*(4481), 453–458.

Utz, C., Degeling, M., Fahl, S., Schaub, F. & Holz, T. (2019). (un) informed consent: Studying gdpr consent notices in the field. *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, 973–990.

UX Booth. (2018). *The power and danger of persuasive design*. Retrieved December 11, 2020, from https://www.uxbooth.com/articles/the-power-and-danger-of-persuasive-design/

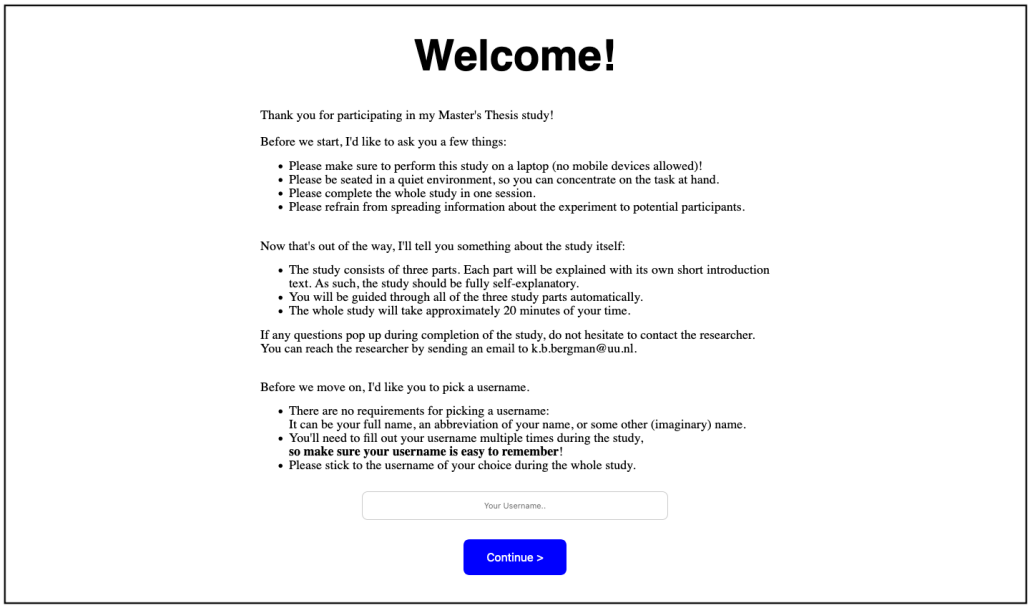
Weinmann, M., Schneider, C. & vom Brocke, J. (2016). Digital nudging. business & information systems engineering 58, 6 (dec. 2016), 433–436.

Whitenton, K. (2018). *The two ux gulfs: Evaluation and execution*. Retrieved December 14, 2020, from https://www.nngroup.com/articles/two-ux-gulfs-evaluation-execution/

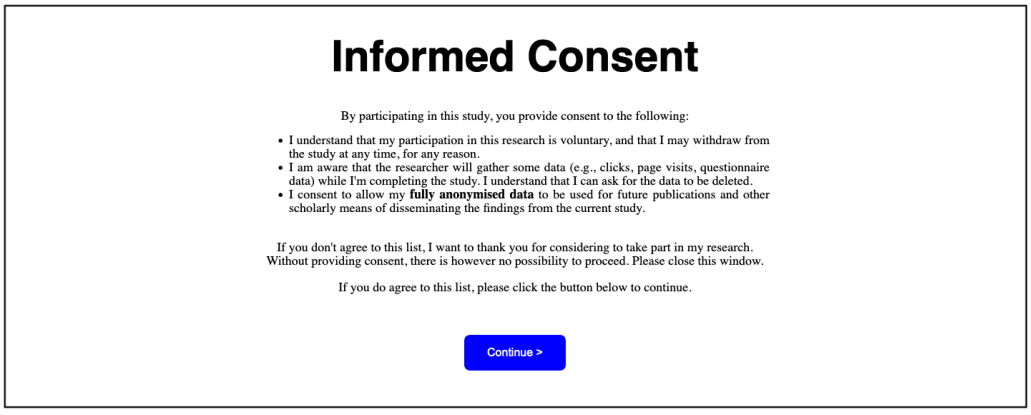
Appendices

A Flow of Study

A.1 Welcome & Instruction



A.2 Consent Form



A.3 General Questionnaire

A.4 Study 1: UX Experiment

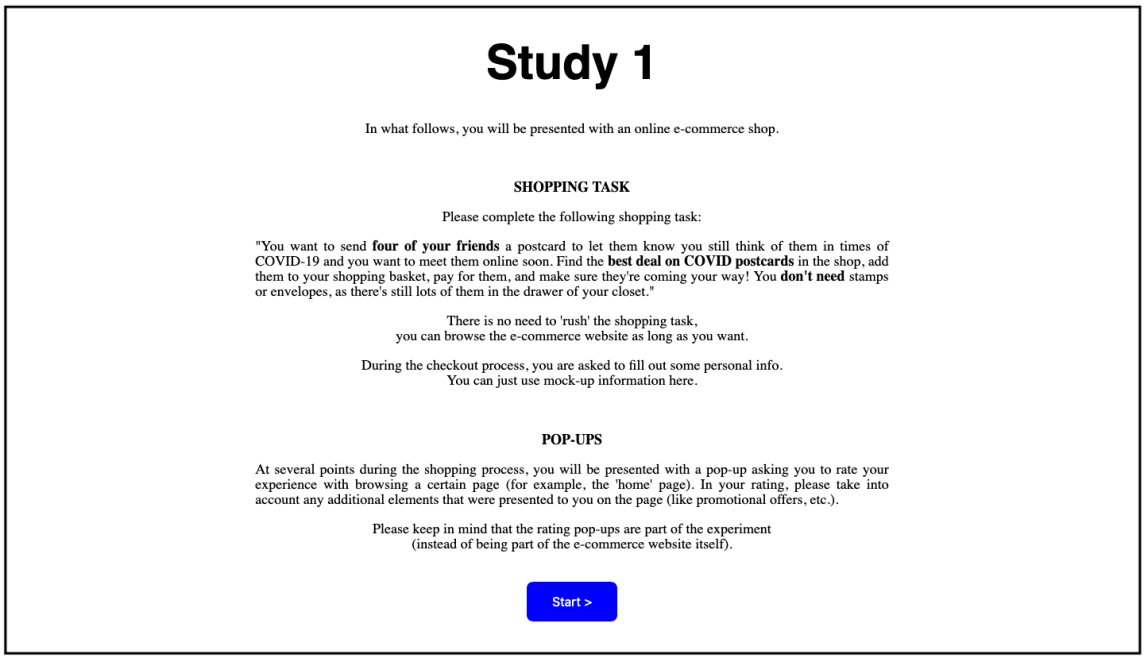
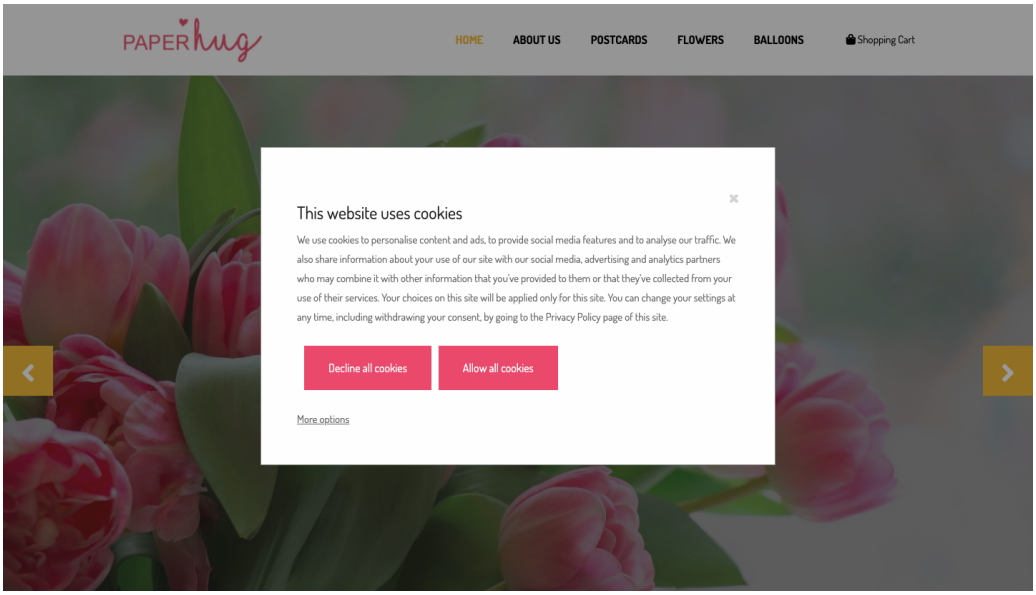
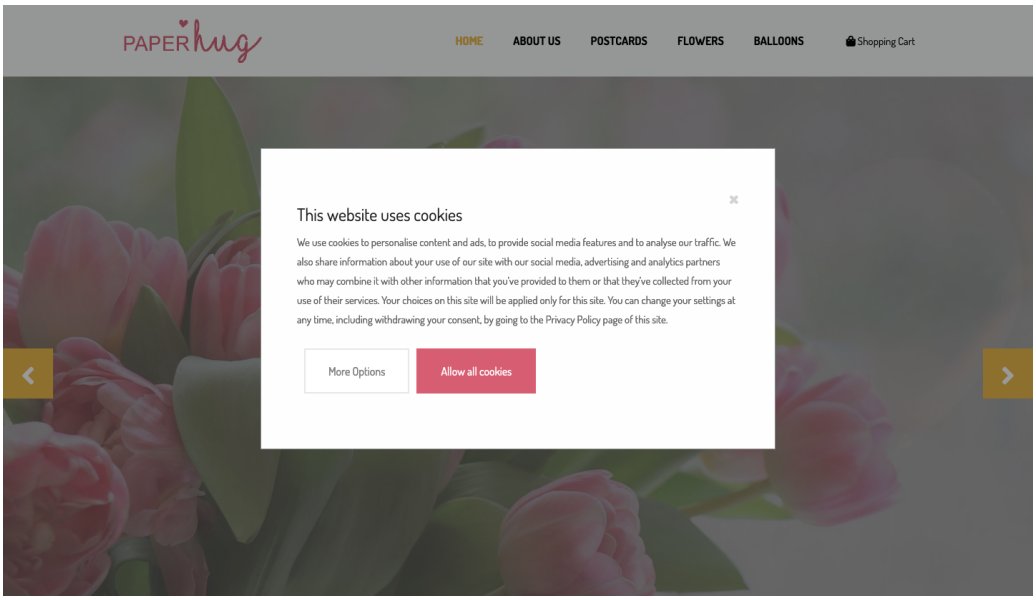


Figure 47: Explanation Page Study 1.

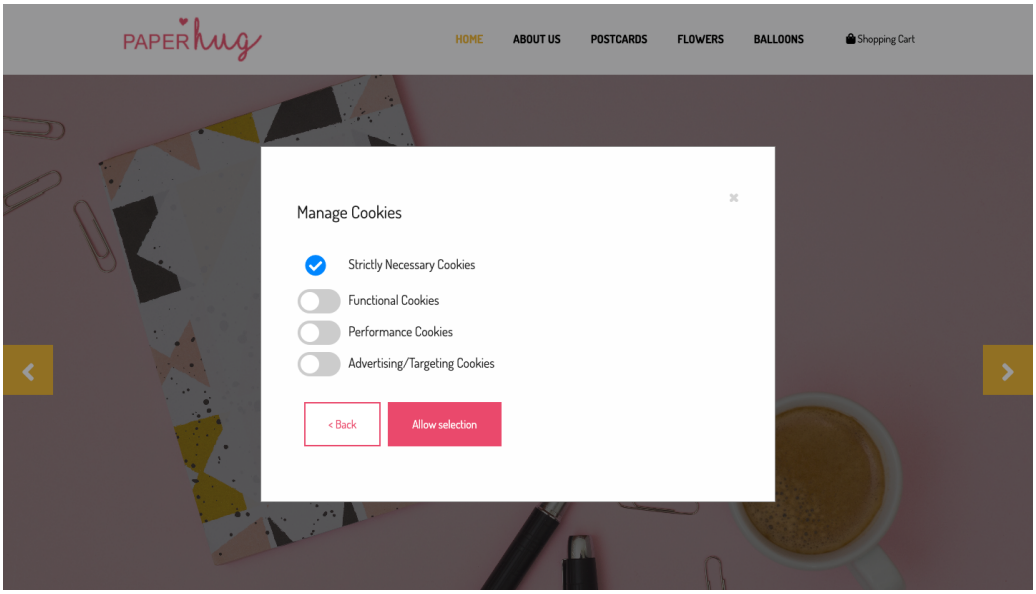


(a) Version A

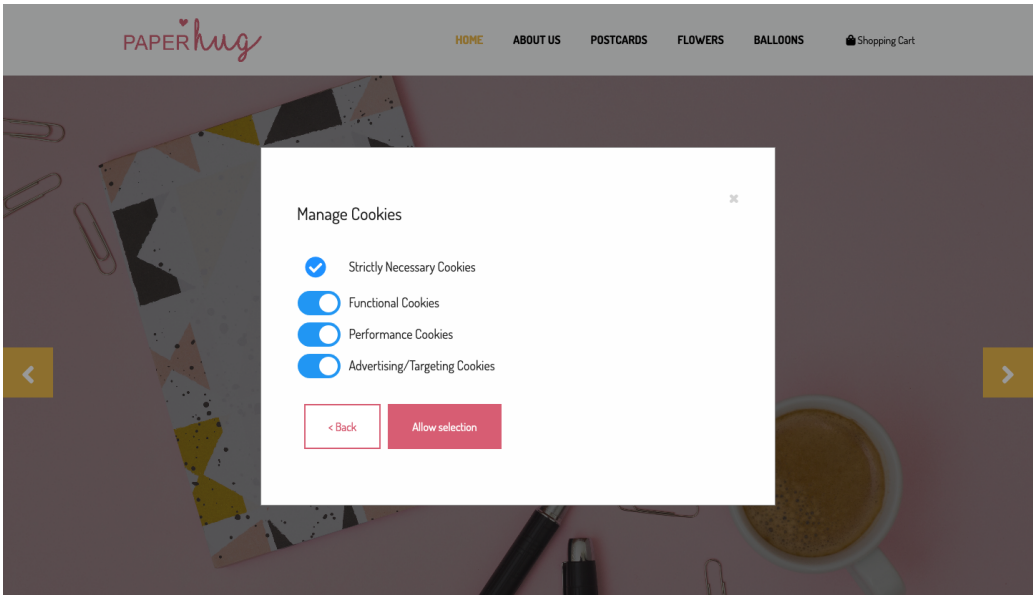


(b) Version B: Hidden Information & False Hierarchy

Figure 48: Home page: Cookie Modal



(a) Version A



(b) Version B: Preselection

Figure 49: Home page: Cookie Modal

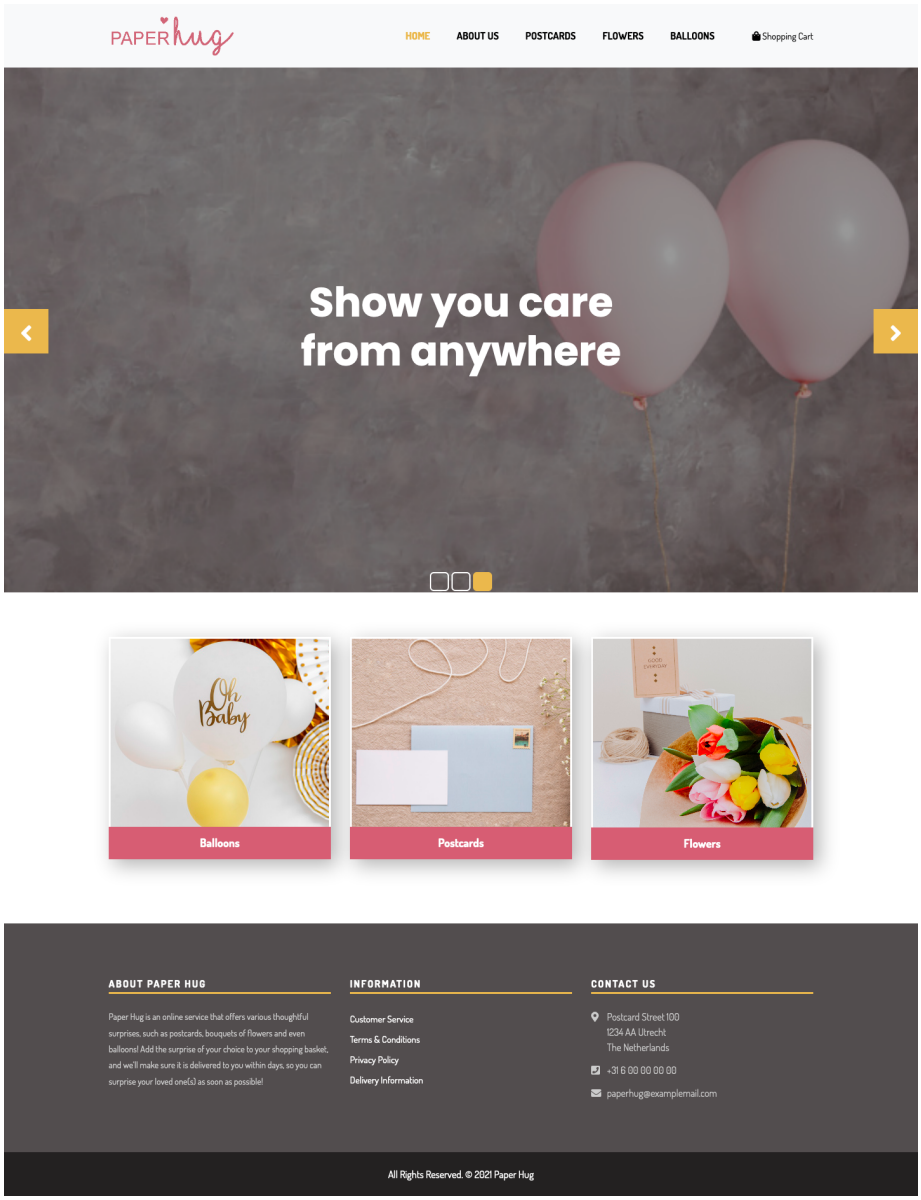
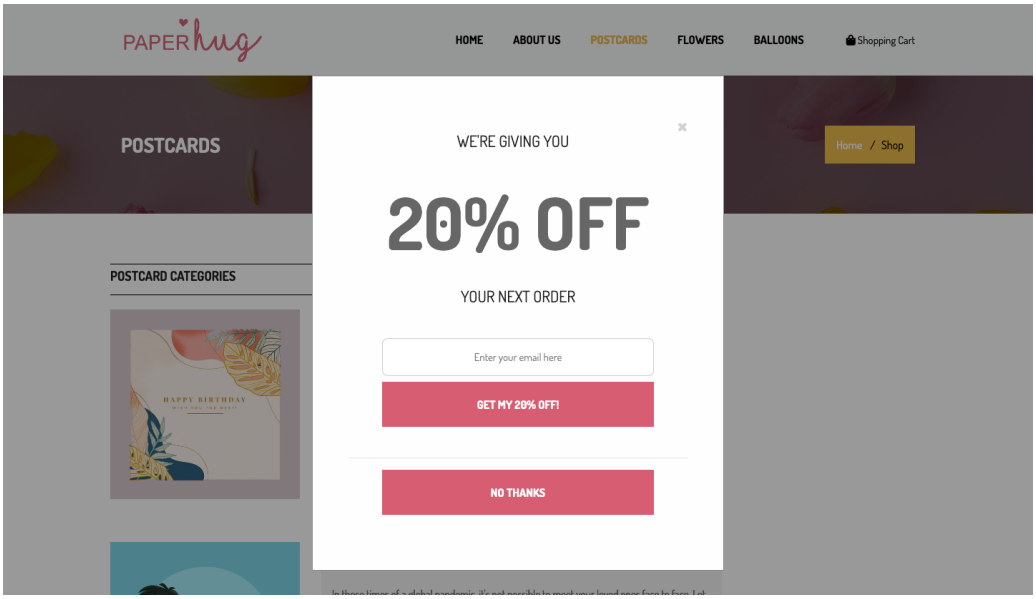
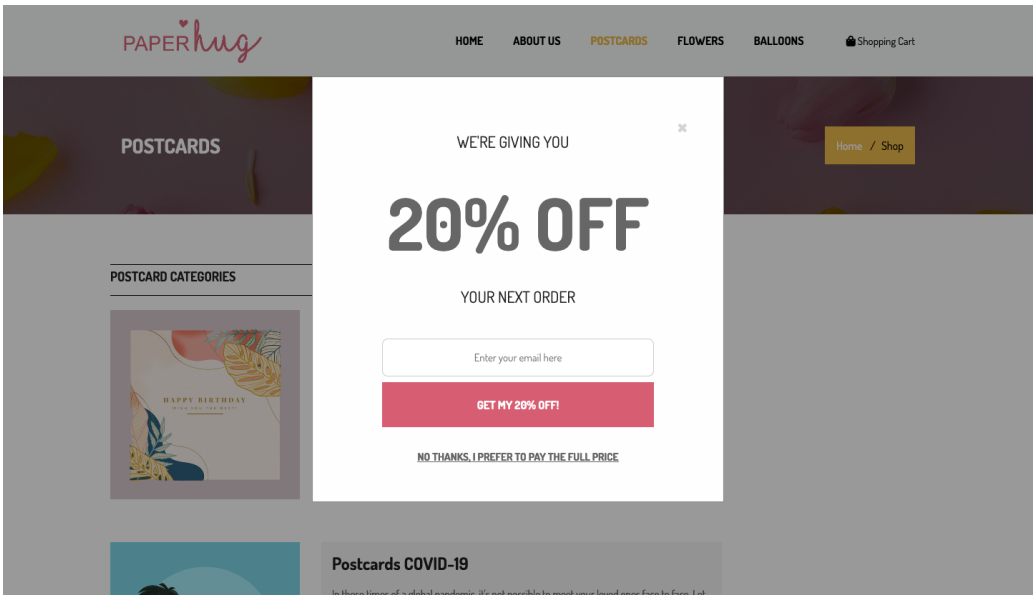


Figure 50: Home page



(a) Version A



(b) Version B: Toying with Emotions & False Hierarchy

Figure 51: Postcard Categories page: Offer pop-up

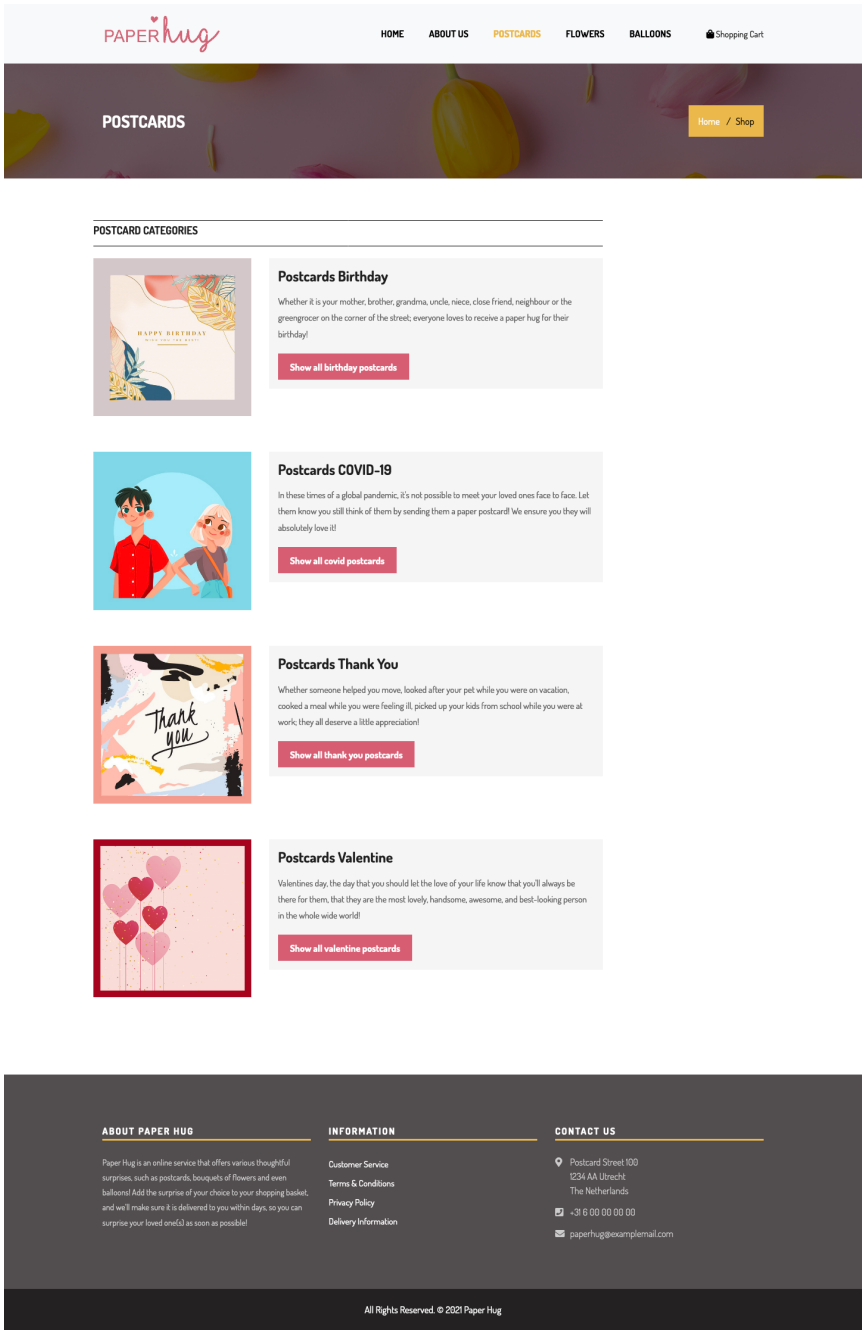
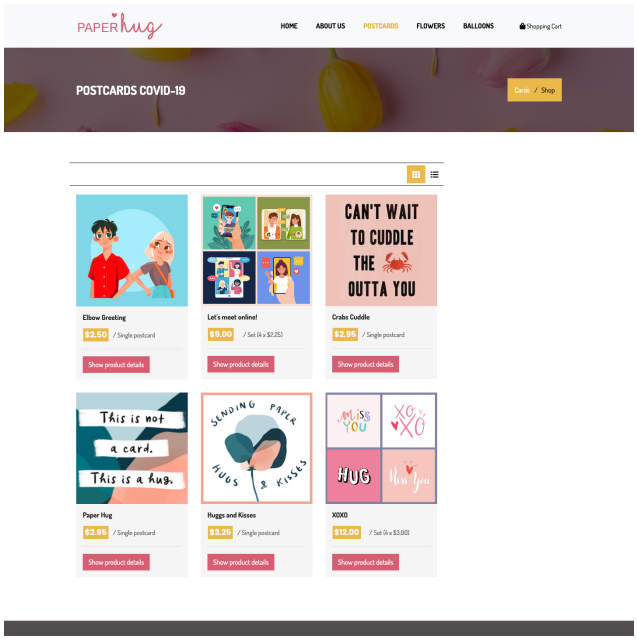
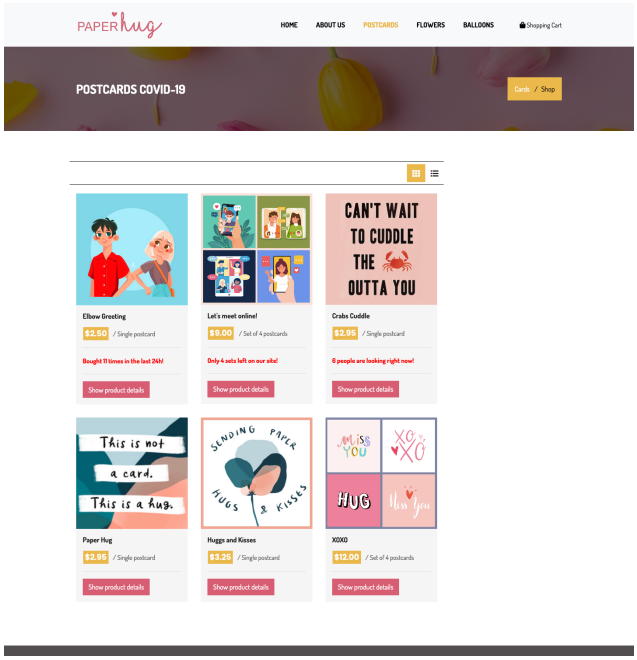


Figure 52: Postcard Categories page

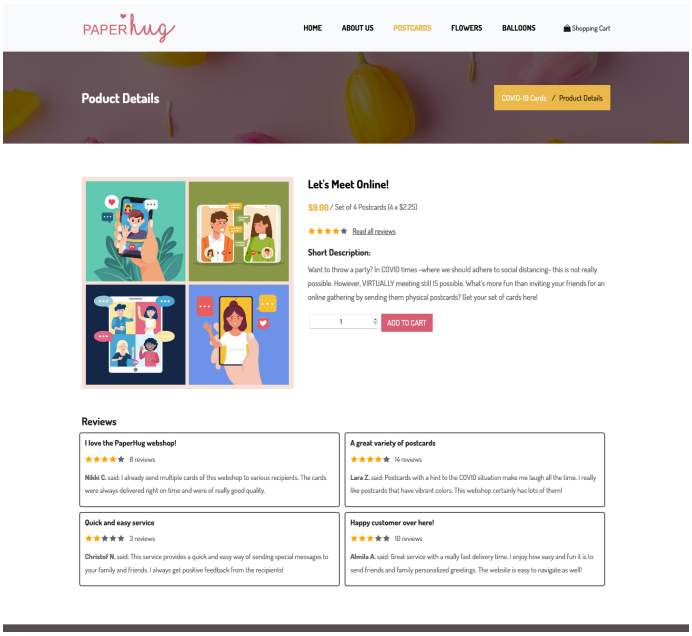


(a) Version A

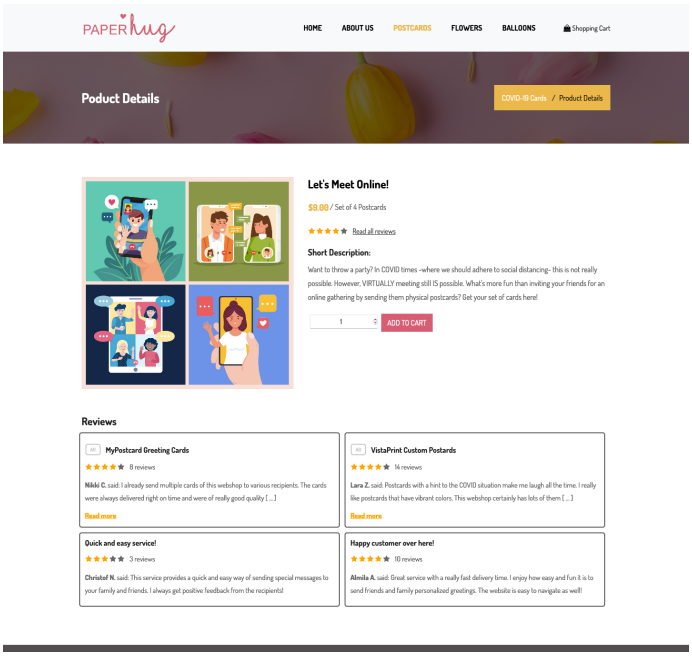


(b) Version B: Toying with Emotions & Price Comparison Prevention

Figure 53: Postcard Overview page

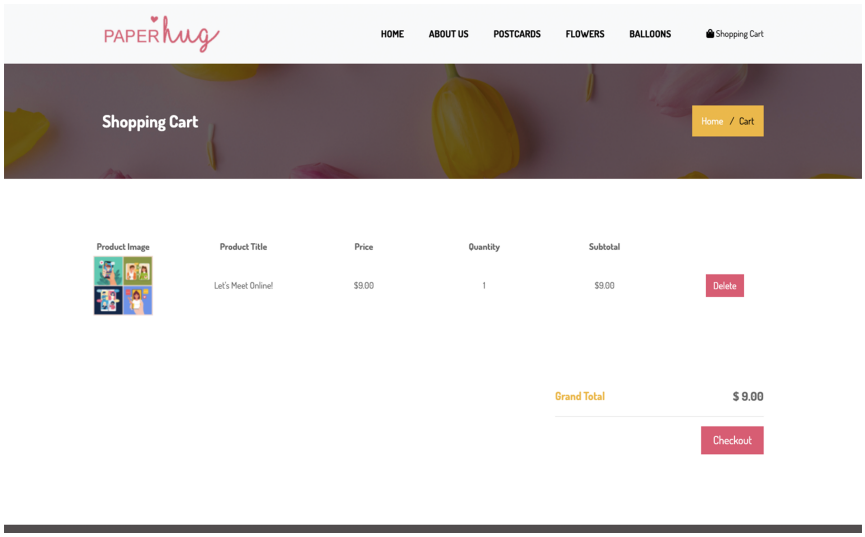


(a) Version A

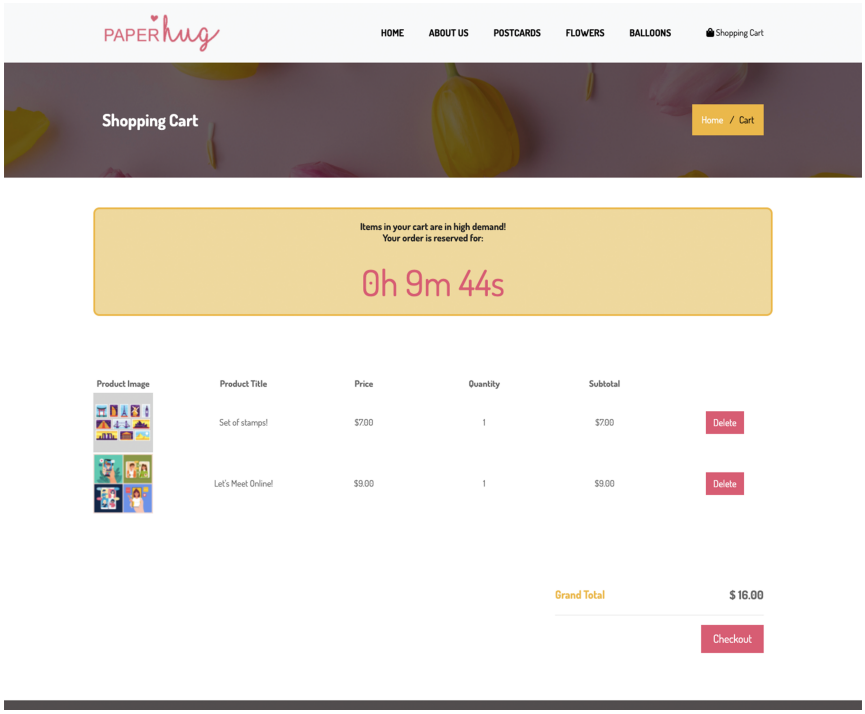


(b) Version B: Disguised Ads

Figure 54: Postcard Details page



(a) Version A



(b) Version B: Toying with Emotions & Sneak into Basket

Figure 55: Shopping Cart page

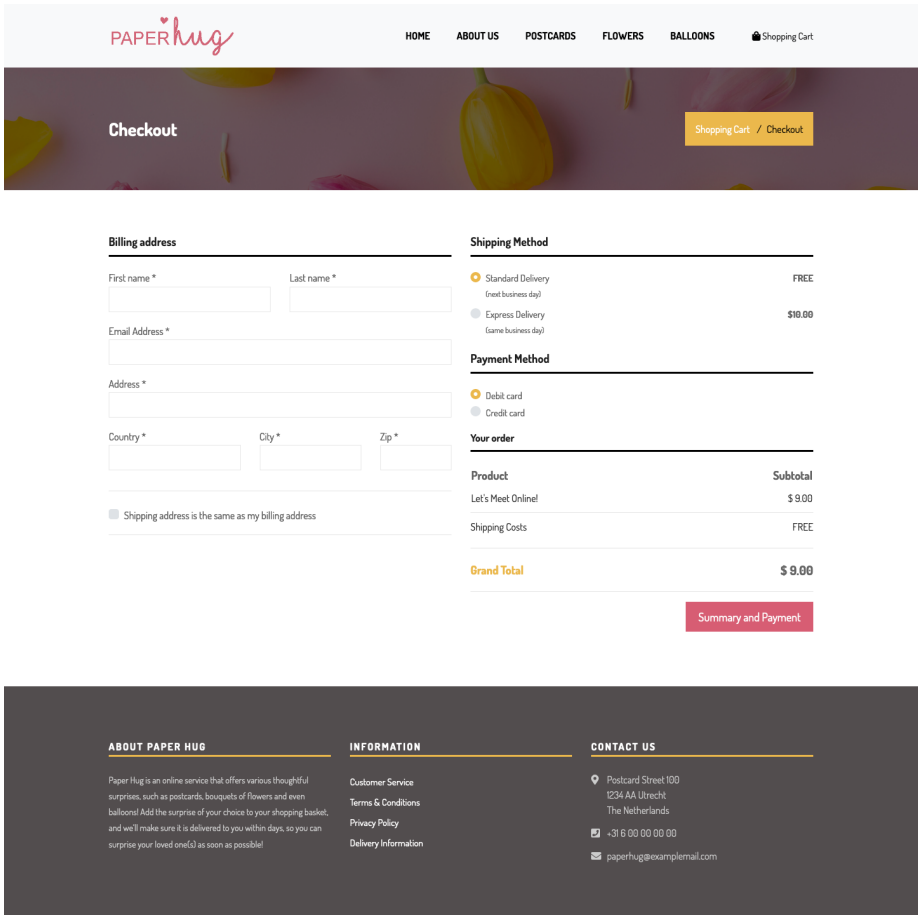
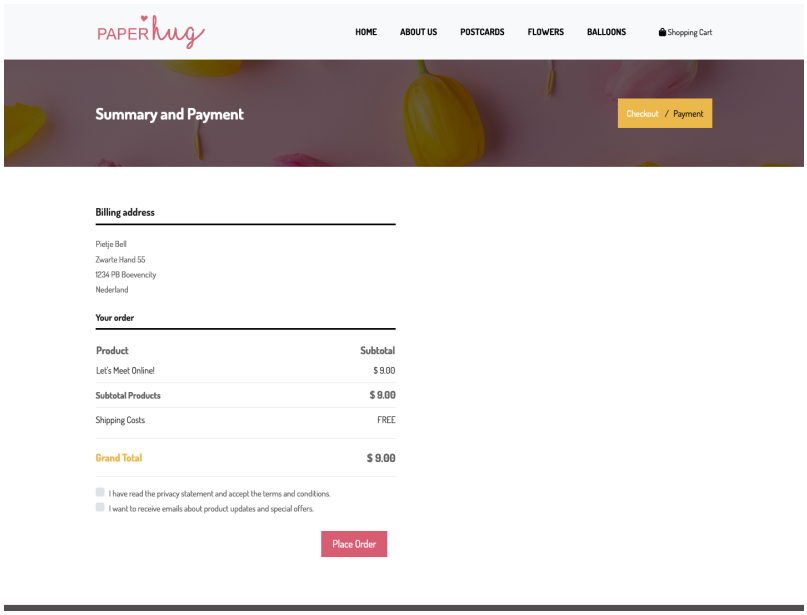
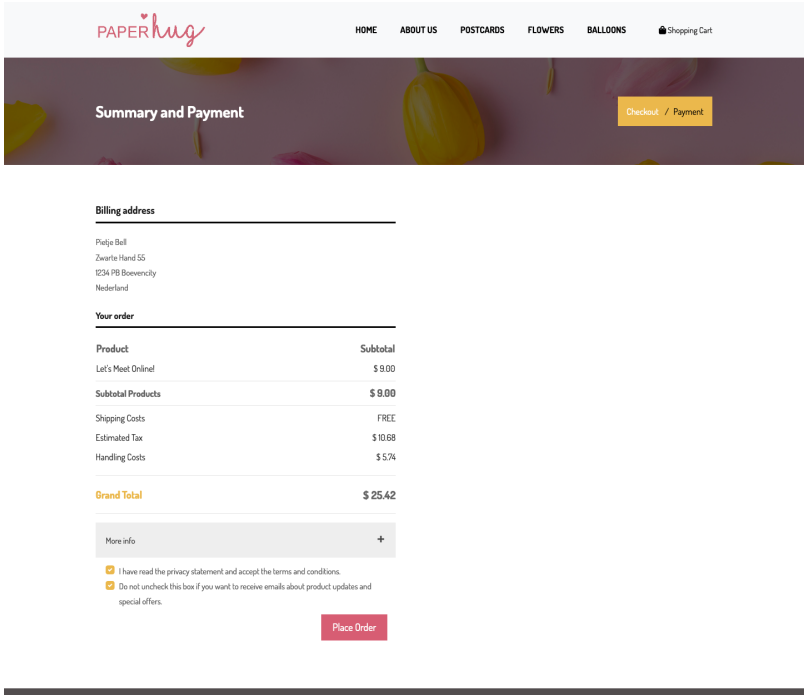


Figure 56: Checkout page



(a) Version A



(b) Version B: Hidden Costs, Hidden Information, Preselection & Trick Questions

Figure 57: Place Order page

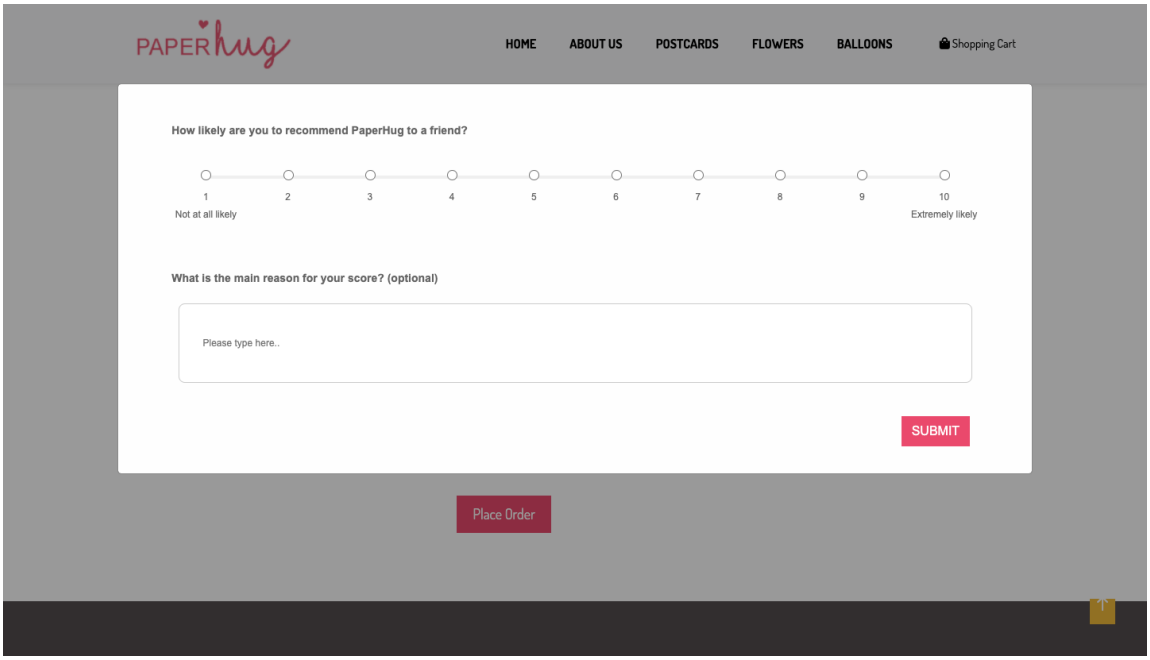


Figure 58: Place Order page: Referral pop-up

A.5 Study 2: SDS Questionnaire

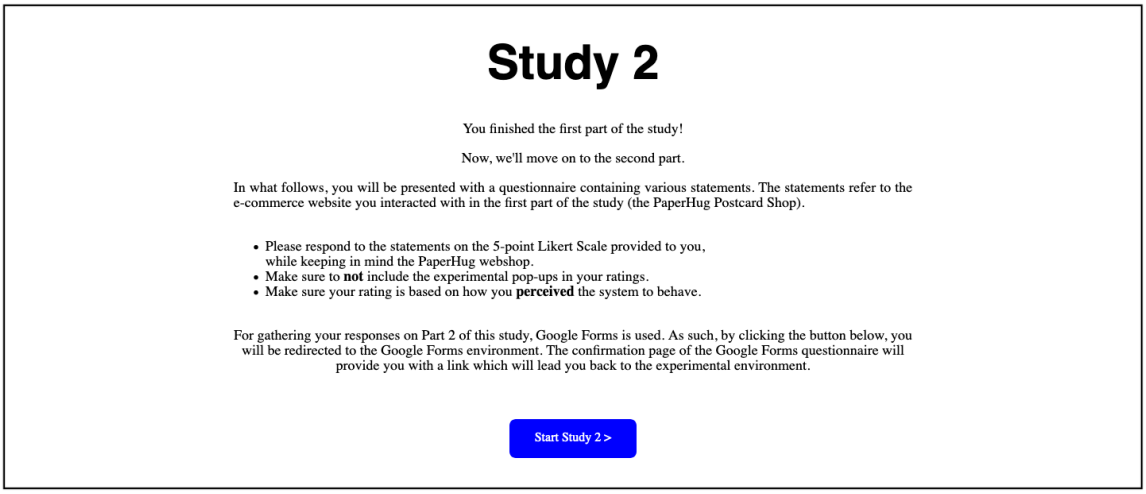
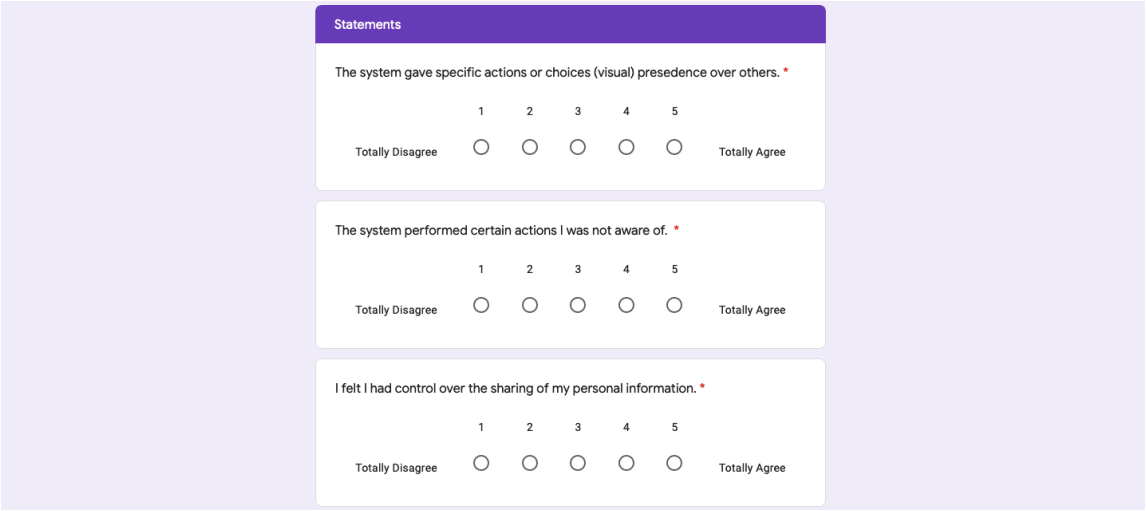
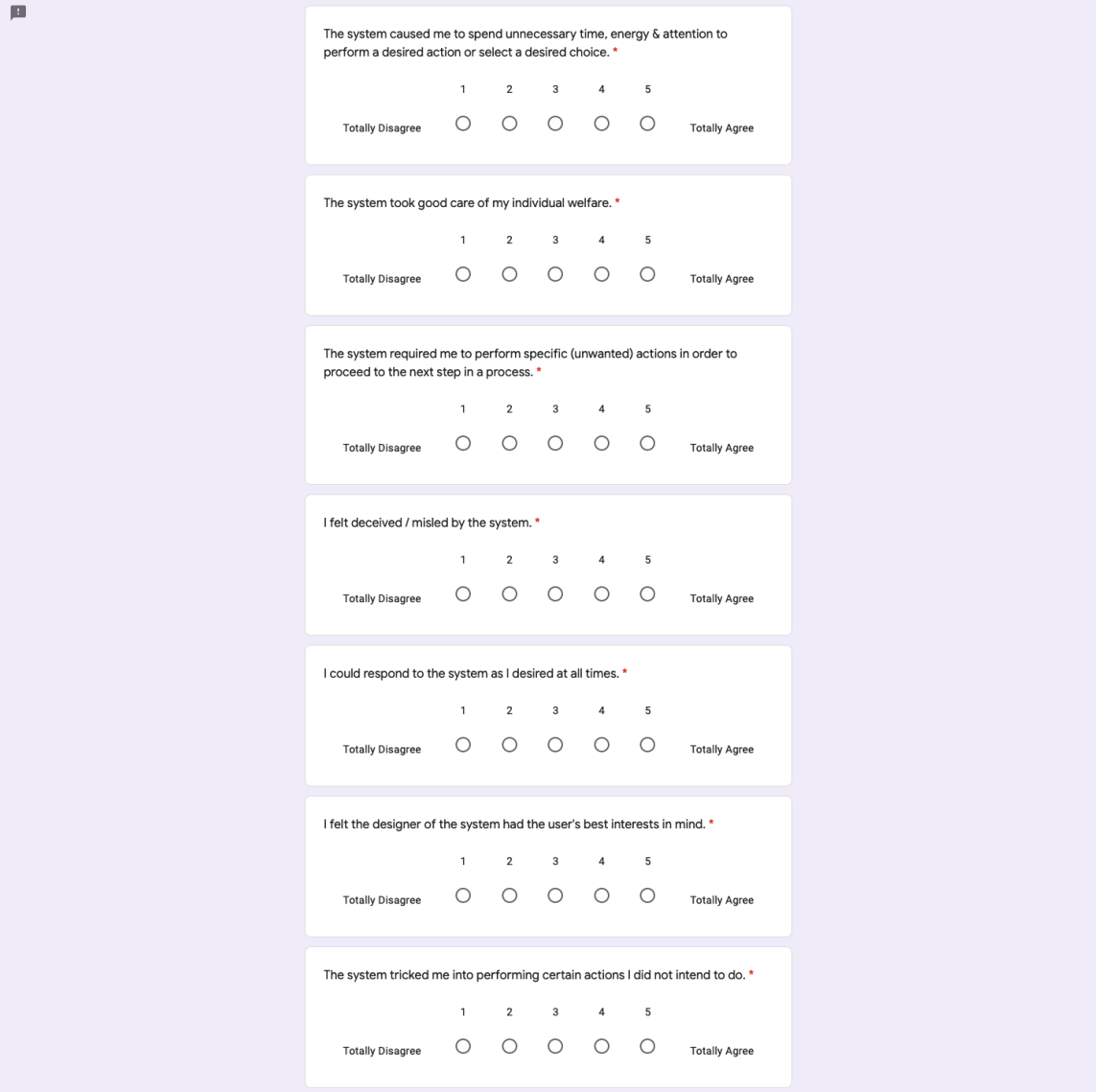
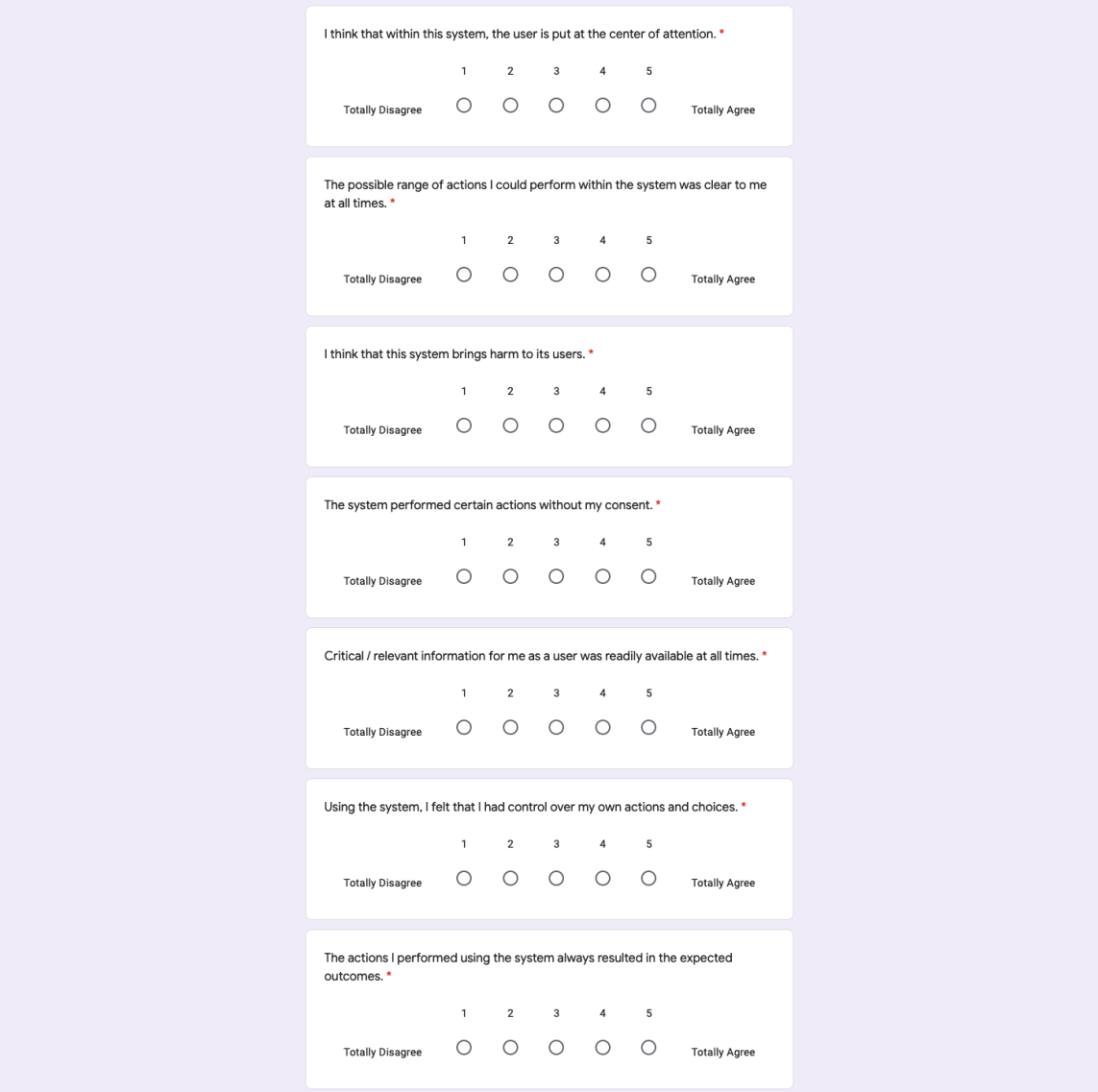
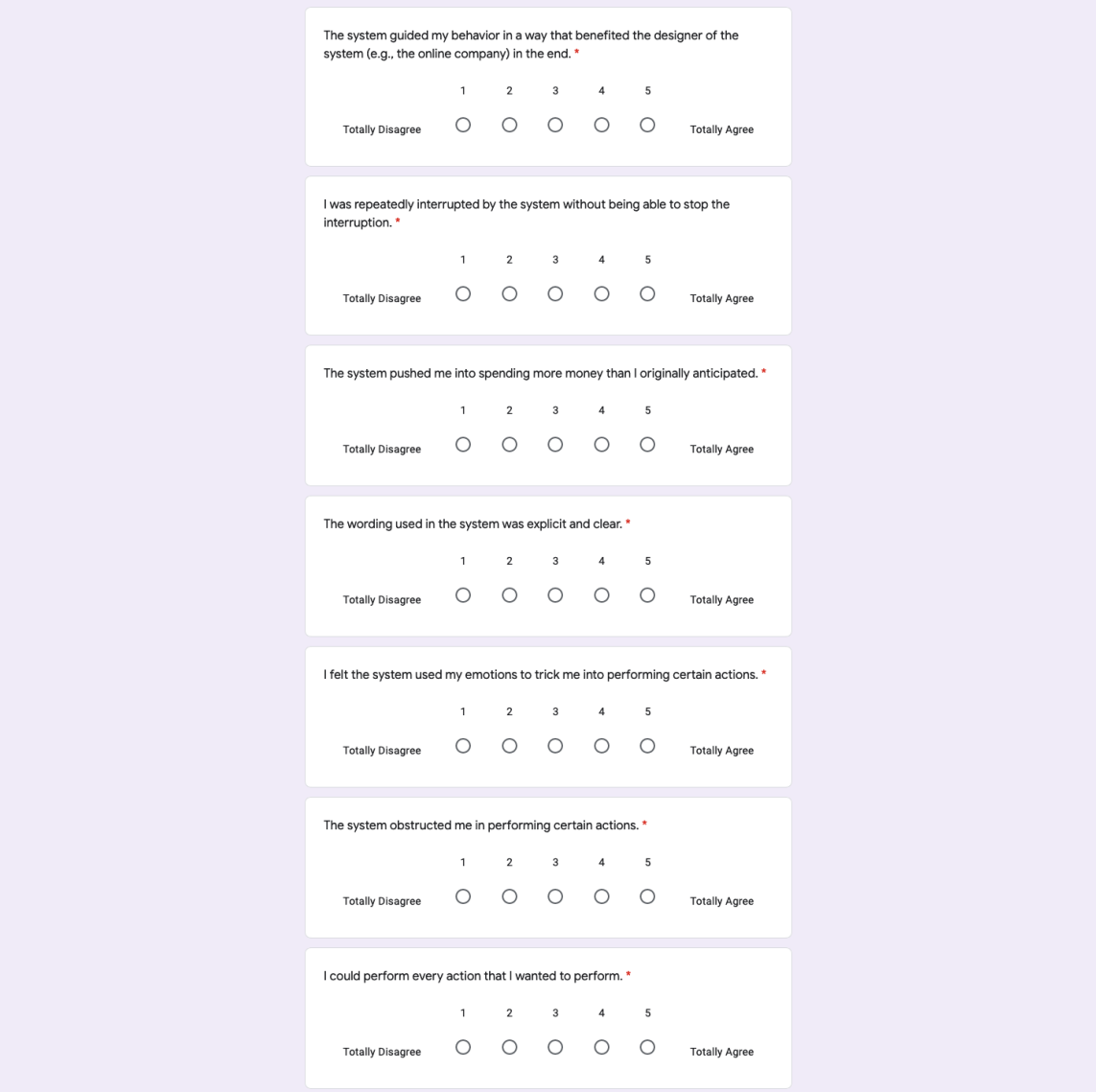


Figure 59: Explanation Page Study 2.




The system caused me to spend unnecessary time, energy & attention to perform a desired action or select a desired choice. *
1 2 3 4 5
Totally Disagree
Totally Agree
The system took good care of my individual welfare. *
1 2 3 4 5
Totally Disagree
Totally Agree
The system required me to perform specific (unwanted) actions in order to proceed to the next step in a process. *
1 2 3 4 5
Totally Disagree
Totally Agree
I felt deceived / misled by the system. *
1 2 3 4 5
Totally Disagree
Totally Agree
I could respond to the system as I desired at all times. *
1 2 3 4 5
Totally Disagree
Totally Agree
I felt the designer of the system had the user's best interests in mind. *
1 2 3 4 5
Totally Disagree
Totally Agree
The system tricked me into performing certain actions I did not intend to do. *
1 2 3 4 5
Totally Disagree
Totally Agree


I think that within this system, the user is put at the center of attention. *
1 2 3 4 5
Totally Disagree
Totally Agree
The possible range of actions I could perform within the system was clear to me at all times. *
1 2 3 4 5
Totally Disagree
Totally Agree
I think that this system brings harm to its users. *
1 2 3 4 5
Totally Disagree
Totally Agree
The system performed certain actions without my consent. *
1 2 3 4 5
Totally Disagree
Totally Agree
Critical / relevant information for me as a user was readily available at all times. *
1 2 3 4 5
Totally Disagree
Totally Agree
Using the system, I felt that I had control over my own actions and choices. *
1 2 3 4 5
Totally Disagree
Totally Agree
The actions I performed using the system always resulted in the expected outcomes. *
1 2 3 4 5
Totally Disagree
Totally Agree


The system guided my behavior in a way that benefited the designer of the system (e.g., the online company) in the end. *
1 2 3 4 5
Totally Disagree
Totally Agree
I was repeatedly interrupted by the system without being able to stop the interruption. *
1 2 3 4 5
Totally Disagree
Totally Agree
The system pushed me into spending more money than I originally anticipated. *
1 2 3 4 5
Totally Disagree
Totally Agree
The wording used in the system was explicit and clear. *
1 2 3 4 5
Totally Disagree
Totally Agree
I felt the system used my emotions to trick me into performing certain actions. *
1 2 3 4 5
Totally Disagree
Totally Agree
The system obstructed me in performing certain actions. *
1 2 3 4 5
Totally Disagree
Totally Agree
I could perform every action that I wanted to perform. *
1 2 3 4 5
Totally Disagree
Totally Agree

A.6 Study 3: DPDS Questionnaire

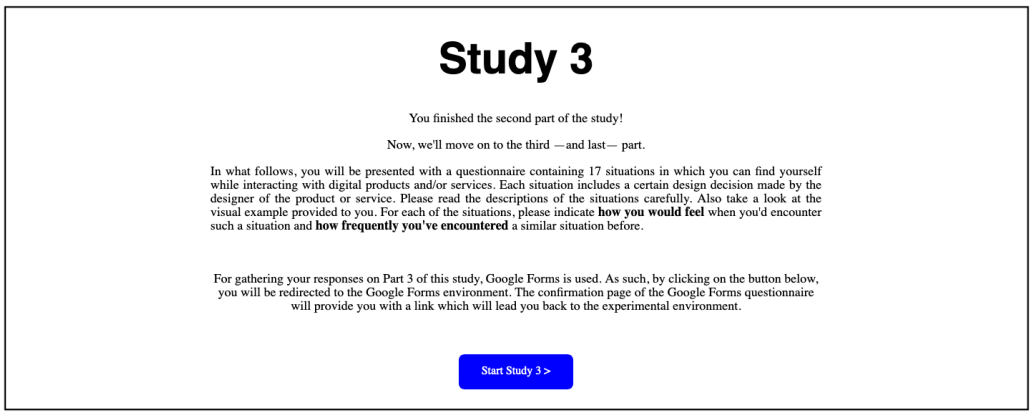


Figure 60: Explanation Page Study 3.

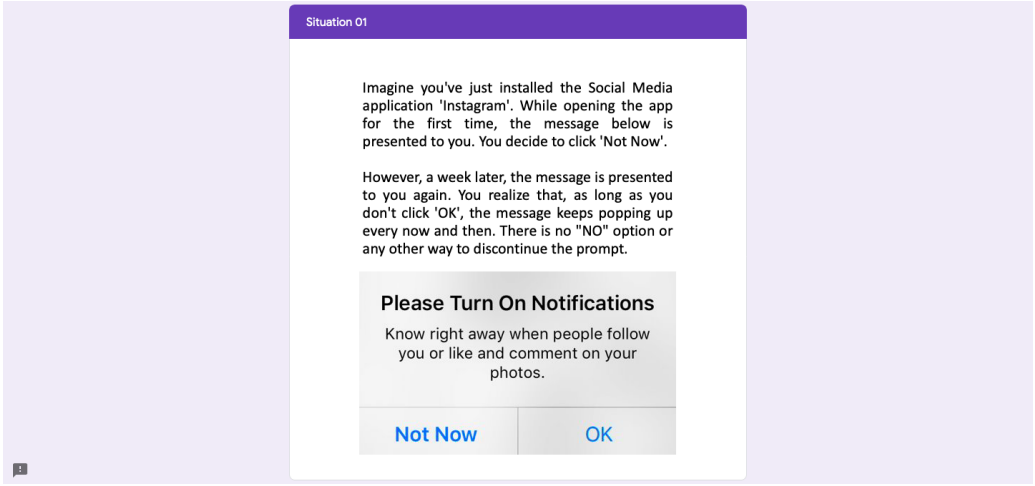


Figure 61: Situation 01: Nagging

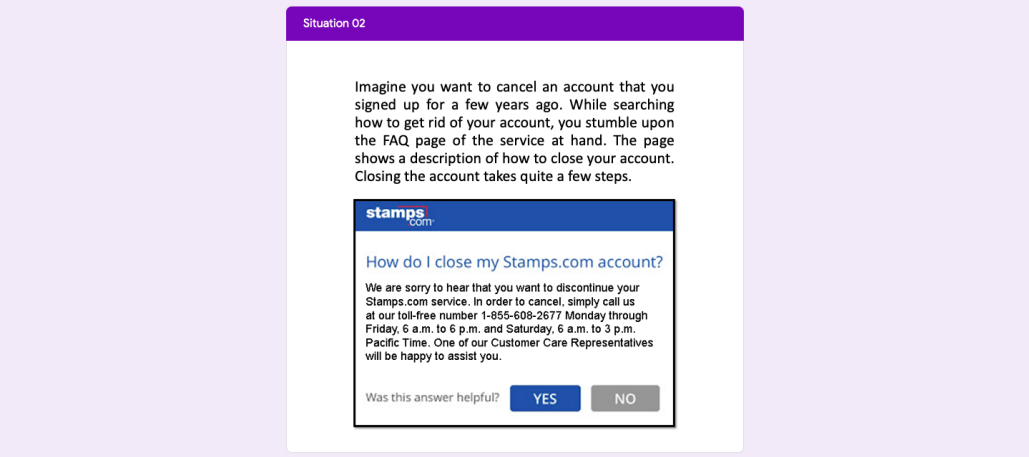


Figure 62: Situation 02: Roach Motel

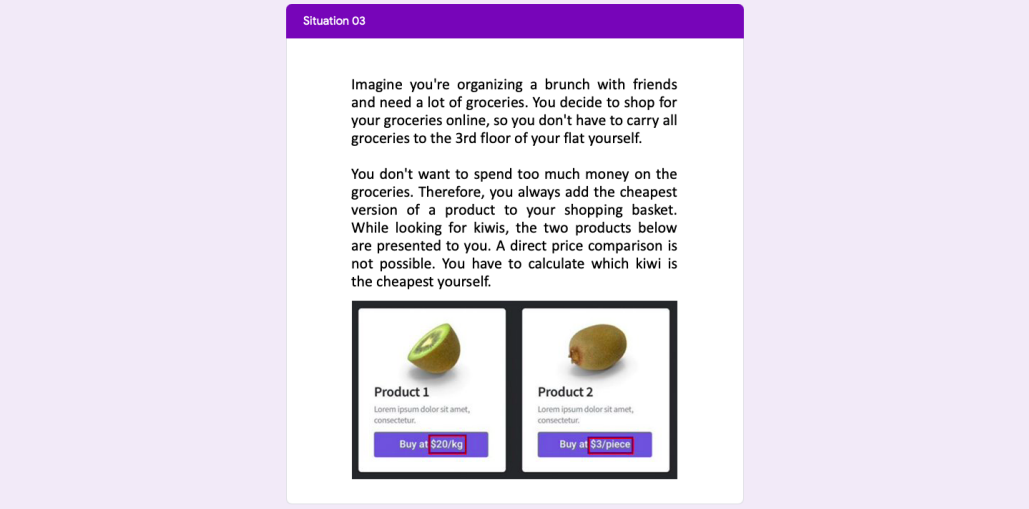


Figure 63: Situation 03: Price Comparison Prevention

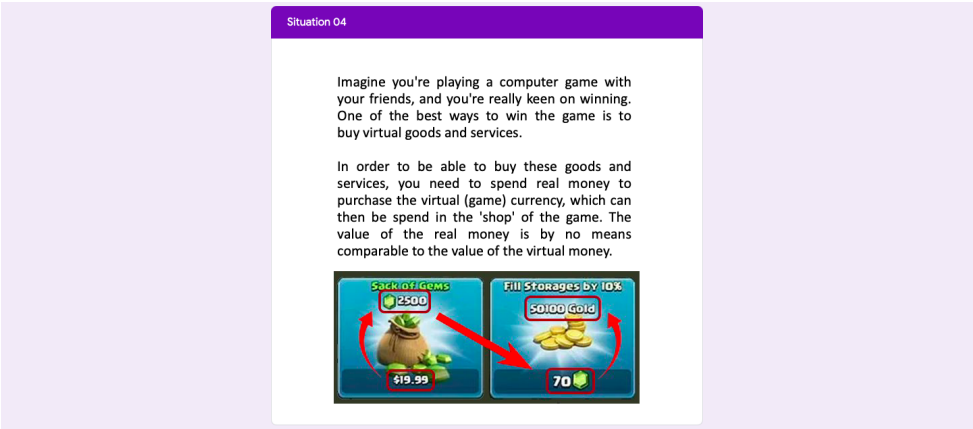


Figure 64: Situation 04: Intermediate Currency

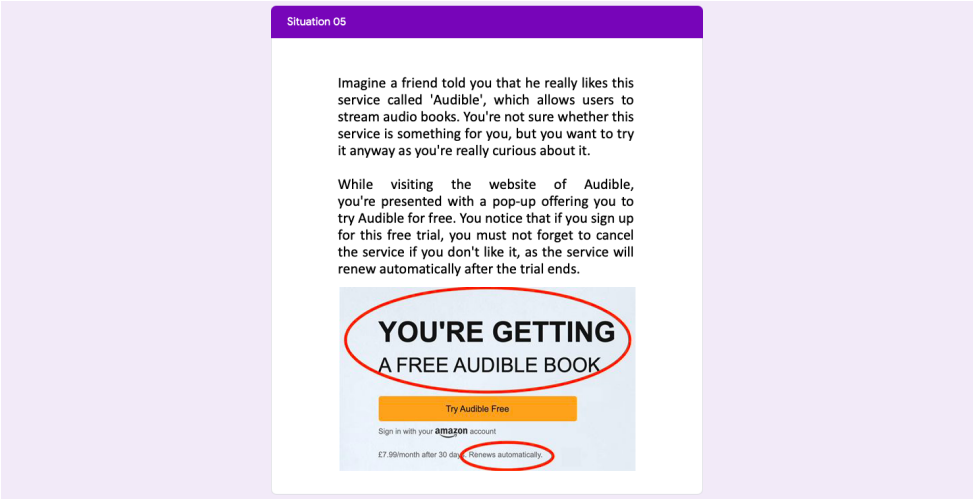


Figure 65: Situation 05: Forced Continuity

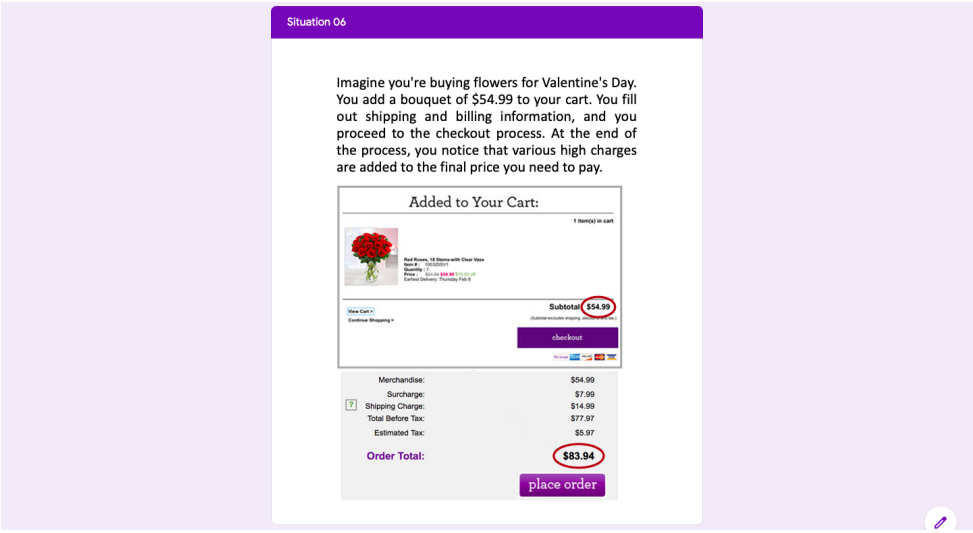


Figure 66: Situation 06: Hidden Costs

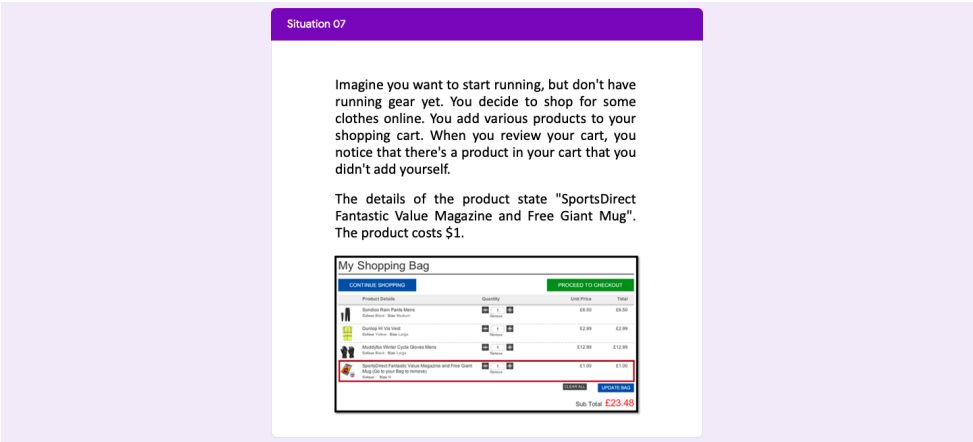


Figure 67: Situation 07: Sneak into Basket

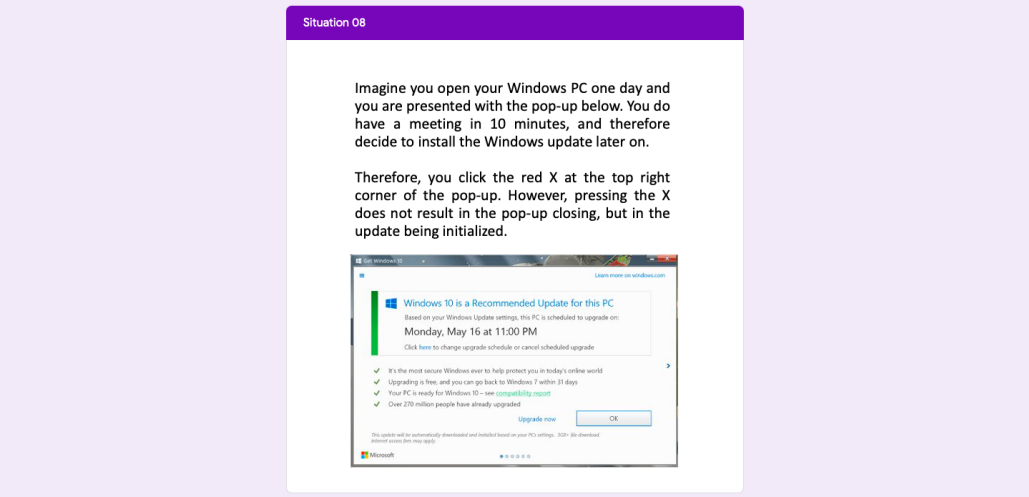


Figure 68: Situation 08: Bait and Switch

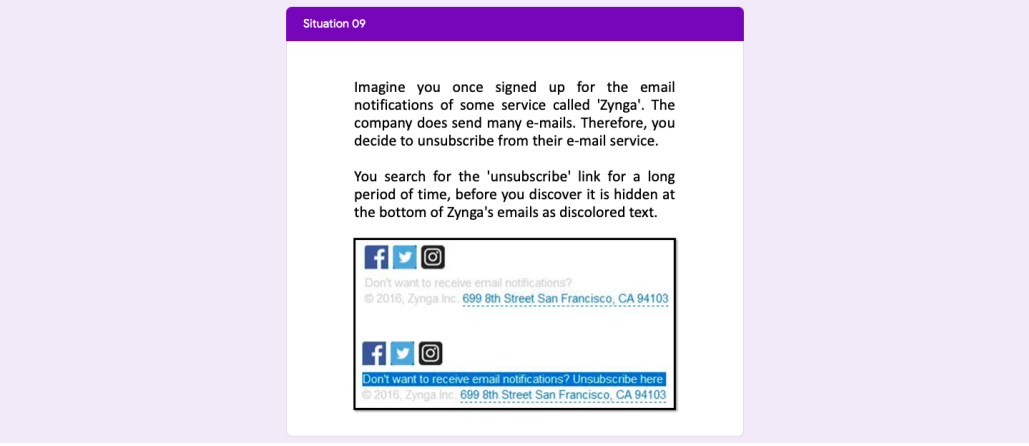


Figure 69: Situation 09: Hidden Information

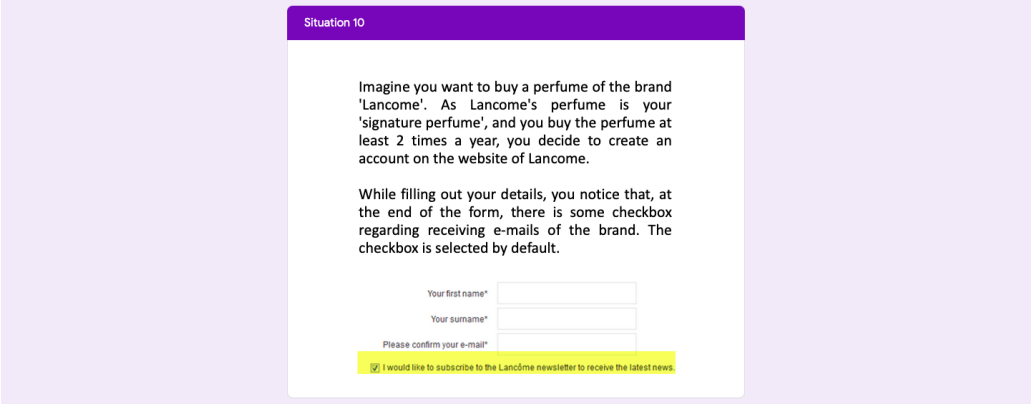


Figure 70: Situation 10: Preselection

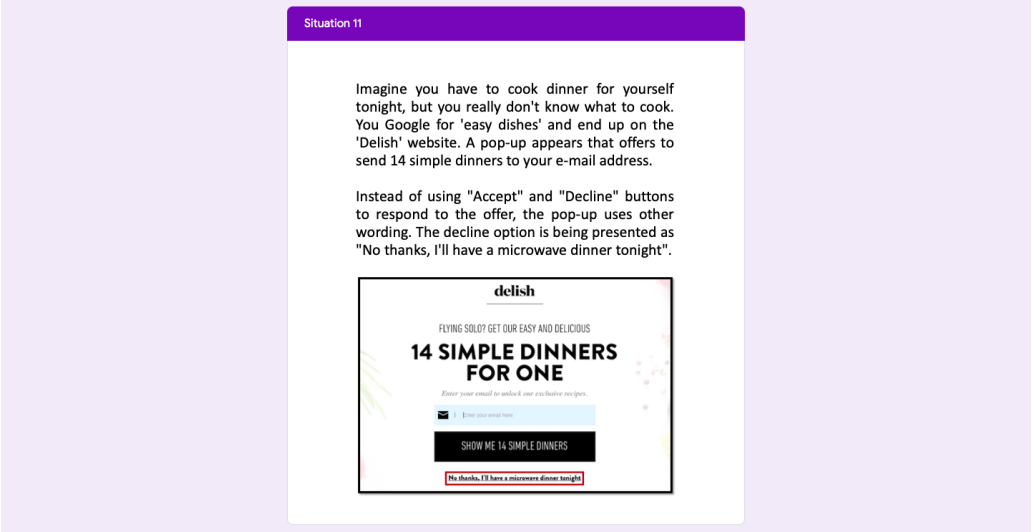


Figure 71: Situation 11: Toying with Emotions

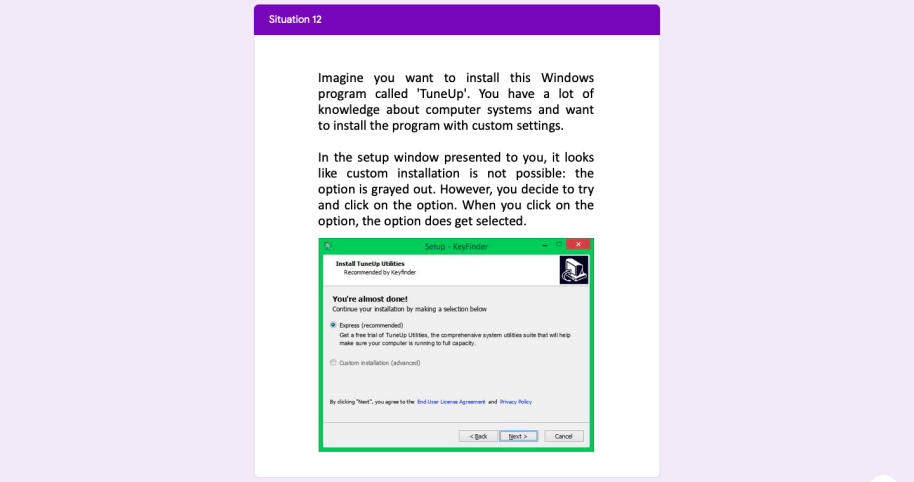


Figure 72: Situation 12: False Hierarchy

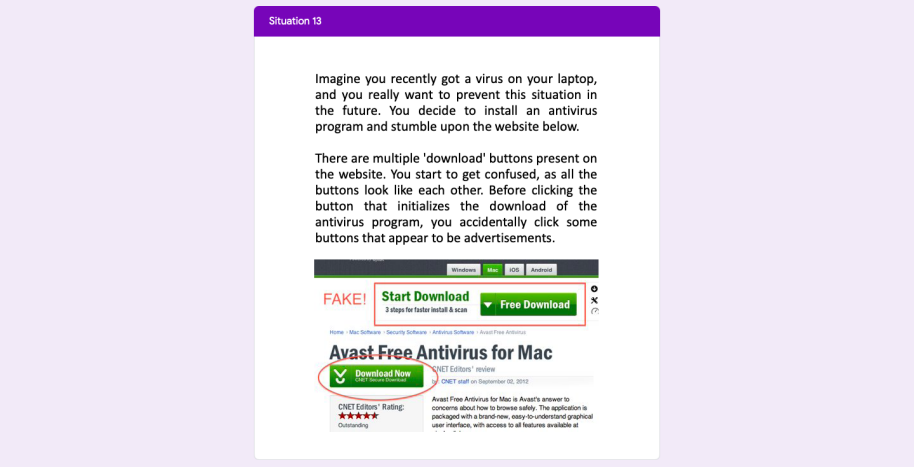


Figure 73: Situation 13: Disguised Ads

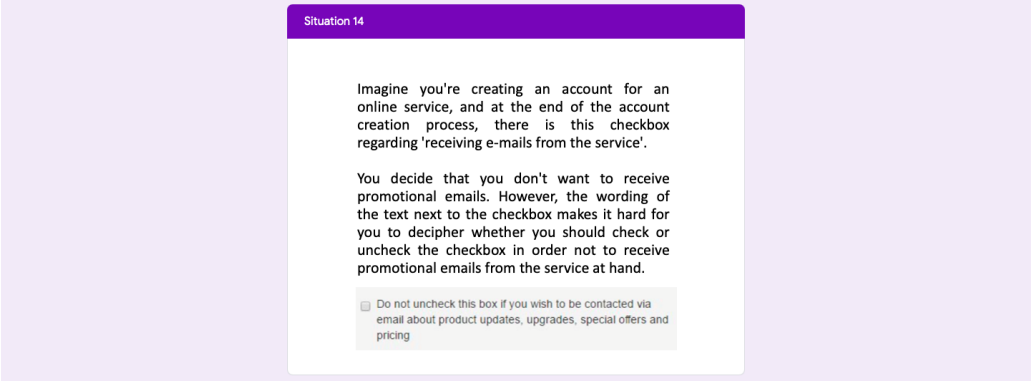


Figure 74: Situation 14: Trick Questions

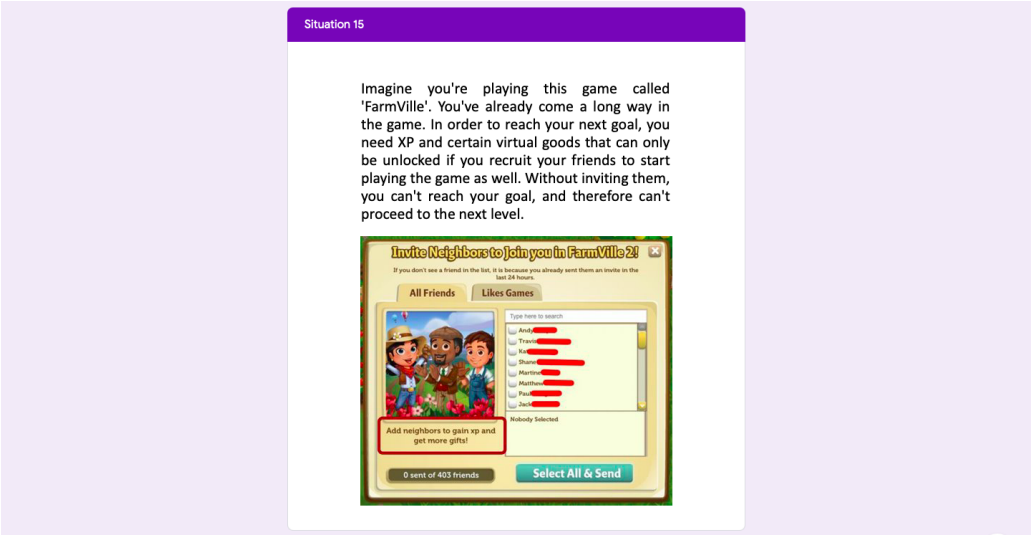


Figure 75: Situation 15: Social Pyramid

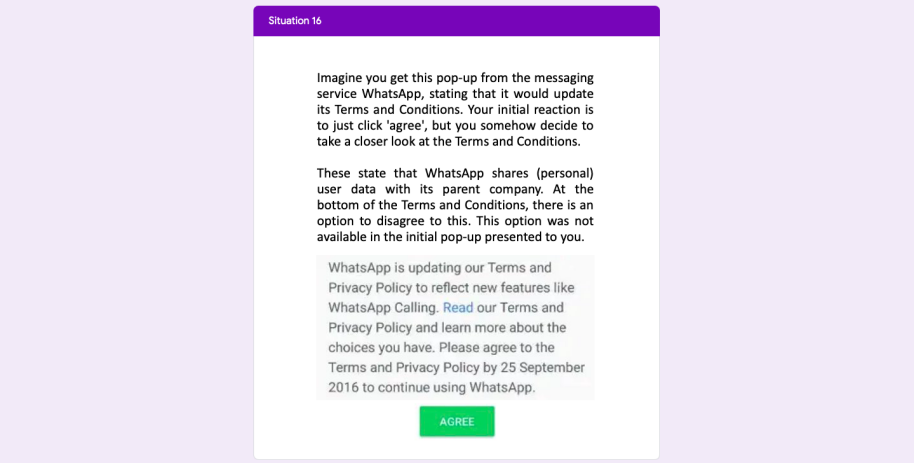


Figure 76: Situation 16: Privacy Zuckering

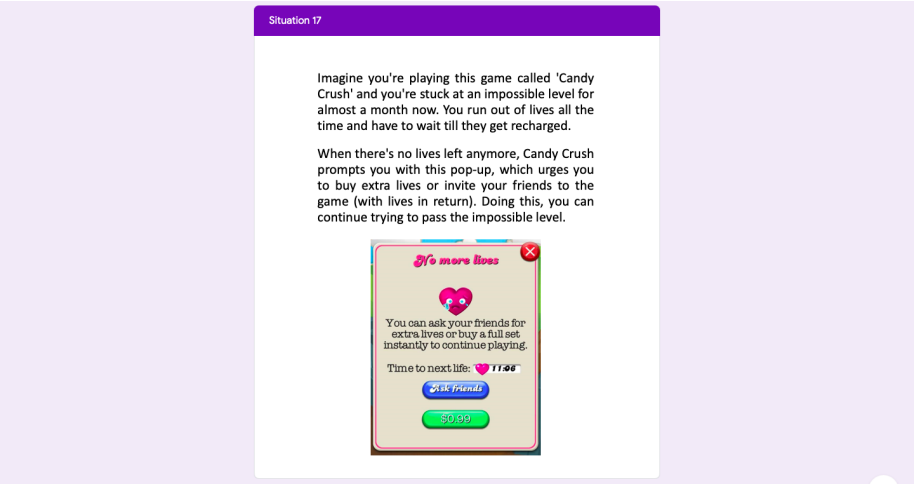


Figure 77: Situation 17: Gamification

A.7 Concluding Remarks

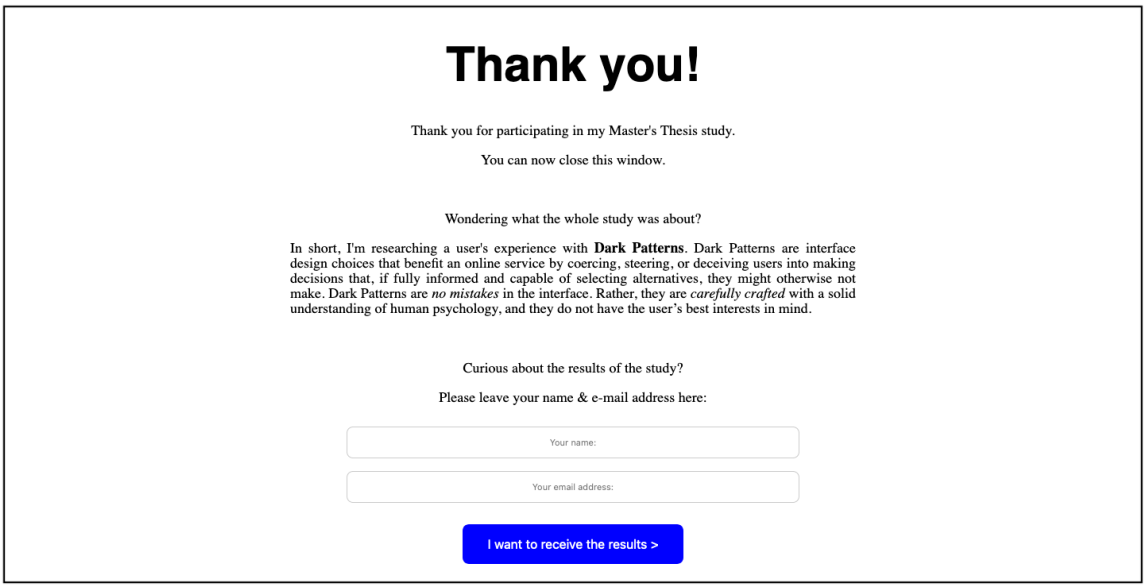


Figure 78: Thank you page.

B Word of Thanks

Writing this Master's Thesis would not have been possible with the help of others.
Therefore, I would like to thank a few people.

First of all, I would like to thank Christof van Nimwegen, which whom I started this journey. He was the one who sparked my interest in Dark Patterns. His creative ideas and enthusiasm really motivated me. Every meeting was filled with coffee, laughter and many new insights. It looked like we could not finish this journey together due to some health issues. However, Christof would not have been Christof if he did not do everything within his power to get back into the game!

Second, I want to thank Almila Akdag, who was designated as my second supervisor. I got lucky that she was already really involved in the process when Christof decided to take a few steps back. Without doubt, Almila took over the job of 'first supervisor' for a while. I am forever grateful that she did, as she fulfilled this role perfectly. I can imagine it not being easy to catch-up with all the (ambitious) ideas Christof and I had in mind, but you really pulled it off -and you slowed me down where necessary ;)!

Third, I want to thank my boyfriend, Rogier Simons, for his endless support. Our many conversations led me to various new insights. He was always there for me when my coding skills lacked, when I could not come up with the English word for 'Dutch term X', or when I did not see how to proceed with certain aspects of my thesis. Working from home was no problem at all, as long as we shared a desk.

Next, I want to thank my family, who were always interested in my thesis process, and really helped me out when I was recruiting participants. Thanks to my parents (and all of their friends), the mean age of my sample really went up: creating a more realistic image of the population ;).

Lara and Rob, thank you for taking the time to provide really in depth feedback on the first version of my Master's Thesis study. Without you, my explanation texts would still have been cluttered with grammatical errors, and my research material would not have been so well thought out.

Last of all, I want to thank all participants that participated in my study, and the many enthusiastic responses I received as a result.