How users interact with dark patterns:
researching effects on time spent and gaze patterns with eye tracking

Master Thesis

David Schaareman

Human Computer Interaction

Utrecht University

First Supervisor

Christof van Nimwegen

Second Supervisor

Almila Akdag

2022-2023

Thesis number: HCI-6402089

Utrecht University

Abstract

Dark patterns (also known as deceptive design) are techniques that can be applied in digital user interfaces to steer the behaviour of the user into a certain direction that is beneficial for the owner of the website or app. A lot of research has been done over the last few years on this topic, but not much is yet known about how users look at instances of dark patterns nor how much time they cost or save users. Hence this research investigates user gazing behaviour using eye tracking equipment and simultaneously measures the time users spend looking at dark patterns and how this influences their total time spent on a task. The research consists of an experiment (N=13) in which people completed two tasks in a randomised order. Both tasks had an experimental version with dark patterns and a control version without dark patterns. Each participant saw one task in the control version and one task in the experimental version. Divided over the two tasks seven instances of dark patterns were implemented: *sneak into basket*, *trick question*, *false hierarchy*, *preselection*, *nagging*, *low-stock/high-demand messages*, and *confirmshaming*. All participants also completed a demographics questionnaire and took part in a retrospective think-aloud session.

The results of the experiment showed that participants spent significantly more time on the checkout page in the experimental version of the first task, in which *sneak into basket*, *trick question* and *false hierarchy* were applied, compared to the control condition without dark patterns. For all of these three patterns it was also found – using Area of Interest analysis – that the participants spent more time looking at the deceptive version of the element, compared to the control version. For the other four patterns no such results were found. For the *preselection* pattern the opposite was found: participants spent on average less time looking at the options when one was already selected. This is probably caused by the fact that they were not required to perform an action in this area if they were satisfied with the selected option.

In addition to time based metrics also the *gaze patterns* (the way the gaze of the participants moved over the screen) were analysed. For some types of dark patterns typical gaze patterns could be distinguished. For the *trick question* for example *regressions* (gaze jumping back to a previous word while reading) were seen often.

This research contributes to the further understanding of how dark patterns are perceived by users. It can help designers to weigh the effects they want to achieve with the application of dark patterns against the adverse effects, such as extra time that is needed for the users. On a more theoretical side it offers extra insight in how users interact with and look at dark patterns on websites. This can also help in determining the severity of types of dark patterns, which in turn can help deciding what legislation is needed. Future research can consist of researching more types of patterns with a larger and a more diverse sample in order to be able to draw stronger conclusions.

Table of contents

1. Introduction

User Experience (UX) is the concept that relates to everything that determines how people feel about a product, what their impression is and how enjoyable it is to use it (Sharp, Rogers, & Preece, 2019, p. 13). A good design can improve the UX of a digital system and provide users with a better product. There are however also occasions in which the design of a system is not merely meant to provide users with the best experience. So-called 'dark patterns' are "*instances where designers use their knowledge of human behavior (e.g., psychology) and the desires of end users to implement deceptive functionality that is not in the user's best interest*" (Gray, Kou, Battles, Hoggatt, & Toombs, 2018, p. 1). The ethics of dark patterns are questioned as they can have a negative impact on individuals, as they might spend more money than intended or provide more data than they would want and can have a negative impact on society, as they can for example harm consumer trust (Bongard-Blanchy, et al., 2021).

An example of a dark pattern is *interface interference*, in which the design of an interface steers the behaviour of the user into a certain direction. This is for example the case is the app shown in Figure 1. Here people are asked to sign up when they start using the app, either with Google, Facebook or their email address. It is however not required to sign up, but the 'skip' button is hidden in the right top. With this design people can easily miss it and will sign up even if it is not necessary.

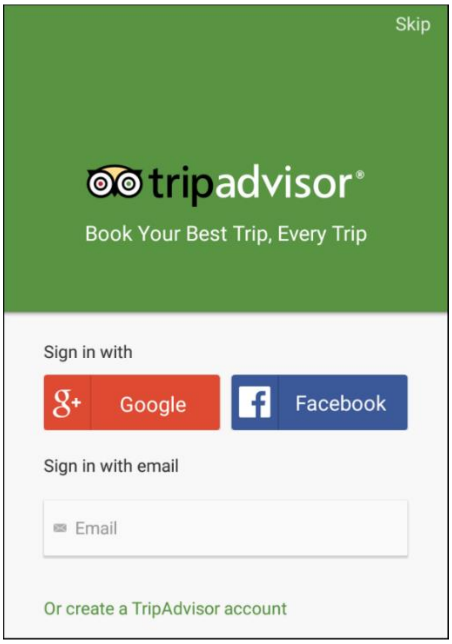

FIGURE 1 EXAMPLE OF A DARK PATTERN: THIS APP CAN BE USED WITHOUT SIGNING IN, BUT THE SKIP BUTTON IS HIDDEN IN THE RIGHT TOP CORNER, SO PEOPLE MIGHT MISS IT. IMAGE FROM (BÖSCH, ERB, KARGL, KOPP, & PFATTHEICHER, 2016).

In recent years a lot of researches on dark patterns have been carried out, that for example have demonstrated that they can be effective in steering the behaviour of users (Nouwens, Liccardi, Veale, Karger, & Kagal, 2020), that they occur quite often (Soe, Nordberg, Guribye, & Slavkovik, 2020) and that they are often not recognised by users (Di Geronimo, Braz, Fregnan, Palomba, & Bacchelli, 2020). As far as known however no research has been done on the question of how the application of dark patterns in a digital service affects the time users need for a certain action. Moreover no known research is yet available on the visual gazing patterns that are provoked by the use of dark patterns.

This research project hence focusses on the time that dark patterns cost users and the eye gazing patterns that can be seen around the interaction with instances of dark patterns. In the project a distinction will be made between local dark patterns and timebound dark patterns. This new way of categorising dark patterns is introduced, as only local dark pattern (i.e. dark patterns that have an effect on the user on a single moment in time) can and will be part of the research (see section 2.3.1. for a full explanation). The research questions for this research will be:

RQ1: How do dark patterns affect the time users need to complete a task?

RQ2: What visual gaze patterns can be seen around applications of dark patterns?

This research can contribute to the understanding of how dark patterns are perceived by users of systems in which dark patterns are used and how they affect them, both time wise and with respect to visual perception. This knowledge can be added to the research that has been done over the past few years in the emerging research field of dark patterns. On a practical side these insights might be useful for owners of systems that use dark patterns as it allows them to better weigh advantages of dark patterns against their potential side effects considering time used by the user and visual influences.

This document starts with a literature review, which first focusses on what is currently known about dark patterns and will subsequently focus on the duration of tasks in a digital environment. At the end of this section the distinction made between local and timebound dark patterns will be explained. The second part contains the methodology for the practical implementation of the research. Next, the results are analysed and subsequently discussed. The document ends with a conclusion and pointers for future research.

Currently there is a shift taking place in which term is being used for the topic concerning this document. Some have suggested that the term 'dark patterns' should be replaced in order to be more inclusive (Sinders, 2022). The term "deceptive design" for example is used instead by Mozilla (Kelly, 2021) and also the one who coined the term 'dark patterns' (Harry Brignull) now uses 'deceptive design' on his website.
The term 'dark patterns' is in academics however (still) far more used than the alternatives, which are currently hardly used in the field. Therefore will this document – aware of the beforementioned considerations – for now also stick to term 'dark patterns'.

2. Literature review

This literature review is split into two parts. The first part will merely focus on dark patterns, and includes what they are, how they work and in which forms and how often they occur. The second part focusses on what is currently known about how users perform a task in a digital environment and how this effects time and visual gaze patterns, as this is relevant background information for RQ1.

2.1. Dark patterns

2.1.1. Introduction to dark patterns

A 'dark pattern' has been defined as "*instances where designers use their knowledge of human behavior (e.g., psychology) and the desires of end users to implement deceptive functionality that is not in the user's best interest*" (Gray, Kou, Battles, Hoggatt, & Toombs, 2018, p. 1). Dark patterns are elements within the design of a digital environment that were purposely implemented to make the users of the system do something that they would not have done themselves otherwise, with the purpose to be beneficial to the system owner or their stakeholders and not necessarily to the user (Cara, 2019).

Dark patterns can be applied in various domains for various reasons. One way they can be applied is to get more personal data from users than they were intending to provide, possibly violating their privacy, as is argued by Bösch, Erb, Kargl, Kopp, & Pfattheicher (2016). Another way dark patterns can be applied is on e-commerce websites, where they can for example be used to persuade users to buy more items or to select a more expensive version of the product (Mathur, et al., 2019). Dark patterns can also be applied in games in the form of 'Dark Game Design Patterns', which can be used to try to make the player spend more time on the game or to pay (more) money to play (Zagal, Björk, & Lewis, 2013). In general Narayanan, Mathur, Chetty and Kshirsagar (2020) summarised dark patterns as being mainly after either users spending more money, giving more data or paying longer attention to a service.

There has been a variety of scientific publications about dark patterns that all have their own defintion for this phenomenon. Mathur, Kshirsagar and Mayer (2021) have published a comparison of 15 academic publications that contain some sort of definition of dark patterns. These are among the various definitions often called 'deceptive', 'misleading' or 'trickery', and their mechanism would 'manipulate users', 'subvert user intent or preferences' and 'trick users'. Part of the defintions include that dark patterns are used to 'benefit the system' or to even 'harm the users' (Mathur, Kshirsagar, & Mayer, 2021). This shows that there is variation between the defitions that exists and that none of them might cover the complete field of dark patterns.

2.1.2. Historical background of dark patterns

Humans are capable of influencing others' behaviour and attitudes. In an article published in 1999 B.J. Fogg introduces the possibility of computers persuading humans, naming it the field of 'captology' (Fogg, 1999). Fogg mentions possibilities of improving ourselves and our society, with respect to safety or health, but also already notes "*But persuasive computers can also be used for destructive purposes (…) [this]*

leads toward manipulation and coercion" (Fogg, 1999), which can be seen as an early mention of the possibility of using digital design for manipulative purposes.

In 2010 Conti & Sobiesk published a paper on malicious interface design, describing an early taxonomy of techniques that can be used for this. Categories in their taxonomy for example include 'confusion', which can for example include asking questions that the user does not understand, or 'forced work', which includes making it hard to uninstall a piece of software (Conti & Sobiesk, 2010).

A phenomenon that is mentioned as the direct predecessor of dark patterns is 'Growth Hacking'. Growth hacking are methods that helps a product grow, using tricks based on design, programming and marketing knowledge (Narayanan, Mathur, Chetty, & Kshirsagar, 2020). Examples of growth hacks include the 'referral program' of online file storage service Dropbox, which gave a user free extra storage space if they invited their friends by providing their email addresses to Dropbox. Another example is free online email service Hotmail adding an advertising sentence for their service at the end of each email sent (McLaughlin, 2014).

From growth hacking the phenomenon of dark patterns has emerged. The term was coined in 2010 by Harry Brignull, who described dark patterns (also named as "deceptive design patterns") as "*tricks used in websites and apps that make you do things that you didn't mean to, like buying or signing up for something.*" (Mathur, et al., 2019, p. 81:4) (Brignull, n.d.). According to Narayanan, Mathur, Chetty & Kshirsagar (2020) one of the key ingredients in the development of dark patterns has been the use of A/B testing. A/B testing is a method that allows to compare two versions of a website or an app. Half of the users sees the regular version (control group) and the other half sees an altered version (experimental group). A variable of interest – such as click rate or visit duration – can be measured and compared between the two versions (Sharp, Rogers, & Preece, 2019, p. 574). This A/B testing technique also proved useful to show how certain small design choices could lead to differences in user behaviour, which could then be used to persuade users to spend more, to provide more data or to stay longer in the app or on the website (Narayanan, Mathur, Chetty, & Kshirsagar, 2020).

In the scientific literature there has been an increasing interest in the phenomenon of dark patterns. The graph in Figure 2 gives an overview of the number of results per year for the query "dark patterns" on the scientific search engine Google Scholar. Until 2015 there were less than 200 results per year, increasing from that moment onwards to almost 1000 results in 2021.

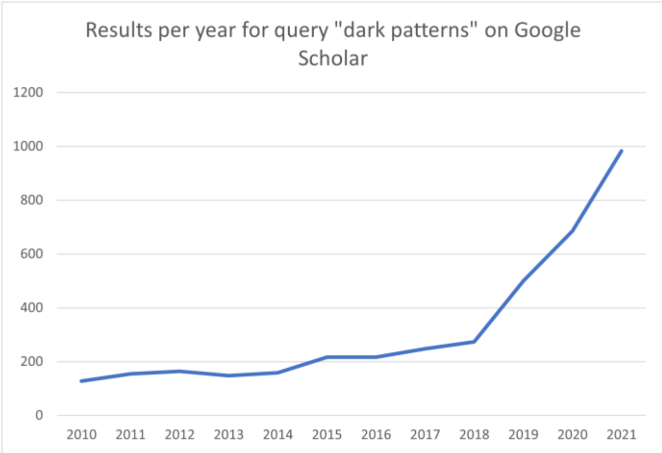

FIGURE 2 GRAPH SHOWING THE NUMBER OF RESULTS PER YEAR FOR THE QUERY "DARK PATTERNS" ON GOOGLE SCHOLAR.

2.1.3. Types of dark patterns

One of the early taxonomies of malicious interface design techniques was published by Conti & Sobiesk (2010). They did then not use the term 'dark patterns' yet, as it was coined only that year by Brignull (see section 2.1.2), who also published an overview of 12 types of dark patterns on his website (Brignull, n.d.). One of the most used taxonomies of dark patterns is the one by Gray, Kou, Battles, Hoggatt & Toombs (2018), which has five main categories of which four have various subcategories. The five main categories are *nagging*, *obstruction*, *sneaking*, *interface interference* and *forced action*. The new preliminary ontology by Gray, Santos and Bielova (2023) also uses these five categories, but also add the new category *social engineering*. The following paragraphs discuss the six categories and the patterns that belong to them. All categories and patterns discussed in this section are shown in table 1.

Nagging. *Nagging* happens when the expected functionality is interrupted, for example when a user is interrupted by a pop-up or distracted by a notice (Gray, Kou, Battles, Hoggatt, & Toombs, 2018). An example of a *nagging* pattern is a supermarket delivery app that repeatedly keeps sending messages to the deliverers to push them to accept lower-paying tasks that they actually might want to reject (Mathur, Kshirsagar, & Mayer, 2021; Eidelson, 2019).

Obstruction. Obstruction is a technique in which a certain action is made more difficult than it needs to be. Patterns that belong in this category are the 'roach motel', 'price comparison prevention' and 'intermediate currency' (Gray, Kou, Battles, Hoggatt, & Toombs, 2018). The roach motel is a pattern in which it is very easy to get for example a subscription, but it is made very hard to unsubscribe (Brignull, n.d.). The price comparison prevention pattern makes it on purpose hard to compare various products, for example by making it hard to copy a product ID, which makes it harder to search for alternatives (Brignull, n.d.; Gray, Kou, Battles, Hoggatt, & Toombs, 2018). The intermediate currency pattern is used when users buy a virtual currency (to be used for in-app or in-game purchases), which might cause the user spending this money differently compared to what they would have done with normal money (Gray, Kou, Battles, Hoggatt, & Toombs, 2018).

Sneaking. Dark patterns in the category sneaking are attempts to make information less easily available to the user to affect their choices (Gray, Kou, Battles, Hoggatt, & Toombs, 2018). The 'forced continuity' pattern is part of this category, which is used to silently start charging users when their free trial ends without a pre-warning (Brignull, n.d.). 'Hidden costs' is a pattern with which users are charged extra costs in the last step of their check-out, such as extra shipping costs (Brignull, n.d.). Also in the domain of e-commerce the 'sneak into basket' pattern can be used, with which a web shop puts an extra item in the shopping basket, for which the user has to opt-out instead of having only items in the basket they have added themselves (Brignull, n.d.). A 'bait and switch' pattern makes an element do something else than expected, for example when the "X"-button does not close the screen but opens a new pop-up instead (Gray, Kou, Battles, Hoggatt, & Toombs, 2018).

Interface interference. Interface interference is a category of dark patterns that uses manipulations in the interface to confuse or affect the user. This can be in the form of 'hidden information', in which relevant information or actions are not immediately visible but hidden in small print, hard to read colours or a long statement. Another form is 'preselection', in which an option that is preferred by the service owner is already selected, such as an option to subscribe to a newsletter. The third option in this category is aesthetic manipulation, in which the layout of an interface is manipulated in a way to affect the user (Gray, Kou, Battles, Hoggatt, & Toombs, 2018). Gray *et al.* (2018) name four specific sub-patterns of this pattern, which are 'toying with emotion', 'false hierarchy', 'disguised ad' and the 'trick question'. Toying with emotion is a dark pattern that tries to evoke emotion, for example when a user has to click a button with "no, I would rather miss fantastic deals" when they do not want to subscribe to a news letter. In a 'false hierarchy' the interface communicates one of the options over another option, such as when one of the options is coloured in grey (Gray, Kou, Battles, Hoggatt, & Toombs, 2018). Disguised ads are advertisements that are designed as regular content with the goal that users click on them. The 'trick question' is a question that is phrased in such a hard or confusing way that it might make a user answer the opposite of what they had intended to answer (Brignull, n.d.).

Forced action. Dark patterns in the category forced action make the user do some (possibly undesired) action in order to continue to some desired point. Instances in this category are the 'social pyramid', 'privacy Zuckering' and 'gamification' (Gray, Kou, Battles, Hoggatt, & Toombs, 2018). The social pyramid dark pattern requires users of a service to invite their friends to also join this service. This can be seen as an extension of the 'Friend spam' pattern defined by Brignull, which means that a service asks access to someone's address book and subsequently sends messages to everyone in it (Brignull, n.d.; Gray, Kou, Battles, Hoggatt, & Toombs, 2018). Privacy Zuckering (named after Facebook founder Mark Zuckerberg) is any form of dark patterns that tries to make a user provide more personal information that they had intended to do (Brignull, n.d.). Finally, gamification is also listed as a dark pattern as it can be used to make users repeatedly use some functionality to earn a certain reward (Gray, Kou, Battles, Hoggatt, & Toombs, 2018).

Social engineering. The category social engineering was introduced in the ontology of Gray, Santos & Bielova (2023). It involves all patterns that are based on social psychological or behavioural economics. It involves for example ‘confirmshaming’, which attempts to give a user a feeling of shame if they would not opt for a certain option (Mathur, et al., 2019). Another example in this category are ‘high-demand and low-stock messages’, which are (sometimes deceptive) texts that state that a product has been sold a lot of times recently, suggesting that it might be unavailable soon, or that a product is already low on stock (Mathur, et al., 2019).

Gray et al. (2018) **category**	**Dark pattern name**	**Mentioned by**				
		Gray et al. (2018)	**Brignull (n.d.)**[1]	**Bösch et al. (2016)**	**Mathur et al. (2019)**	**Gray et al. (2023)**
Nagging		•				Nagging
Obstruction	Roach motel	•	•		•[2]	Obstruction
	Price comparison prevention	•	•			Obstruction
	Intermediate currency	•				Obstruction
Sneaking	Forced continuity	•	•		○[3]	Forced action
	Hidden costs	•	•		•	Sneaking
	Sneak into basket	•	•		•	Sneaking
	Bait and switch	•	•			Sneaking
Interface interference	Hidden information	•				Interf. Inter.
	Preselection	•		○[4]		Interf. Inter.[4]
	Aesthetic manipulation	•			○[5]	
	↳ Toying with emotion	•	○[6]		○[6]	Interf. Inter.[7]
	↳ False hierarchy	•				Interf. Inter.
	↳ Disguised ad	•	•			Sneaking
	↳ Trick question	•	•		•	Interf. Inter.
Forced action	Social pyramid	•	○[8]			Forced action
	Privacy Zuckering	•	•	•		Forced action
	Gamification	•				Forced action
Not in Gray et al.'s (2018) taxonomy	Forced registration			•	•	Forced action
	Hidden legalese stipulations			•		
	Immortal account			•		
	Address Book Leeching			•		Forced action
	Shadow User Profiles			•		

[1] The website of (Brignull, n.d.) was updated in Spring 2023, including a new list of dark patterns. In this table the original list (before the update) is used.
[2] Mathur et al. list ‘Hard to Cancel’ and define this the same way as the ‘Roach Motel’.
[3] Mathur et al. list ‘hidden subscription’, which is related to the ‘forced continuity’ pattern.
[4] Bösch et al. and Gray et al. (2023) mention ‘Bad Defaults’, which is comparable to preselection.
[5] Mathur et al. mention ‘Visual interference’, which is related to ‘aesthetic manipulation’.
[6] Brignull and Mathur et al. list ‘confirmshaming’, which can be seen as a limited form of toying with emotion.
[7] Gray et al. (2023) do not mention Toying with Emotion, but do mention *Emotional or Sensory manipulation* as a meso-level pattern.
[8] Brignull lists ‘friend spam’, which can be seen as a limited form of the social pyramid pattern.

	Countdown timer				•	Social engineer.
	Limited-time Message				•	Social engineer.
	Pressured Selling				•	
	Activity Message				•	Social engineer.
	Testimonials				•	Social engineer.
	Low-stock Message				•	Social engineer.
	High-demand Message				•	Social engineer.
	Privacy maze					Obstruction
	Confirmshaming				•	Social engineer.

TABLE 1 OVERVIEW OF PATTERNS MENTIONED IN THE VARIOUS TAXONOMIES. • = PATTERN MENTIONED; ○ = PATTERN MENTIONED WITH COMPARABLE TERM OR THAT IS SIMILAR. THE LAST COLUMN SHOWS IN WHICH CATEGORY THE PATTERN IS PLACED IN THE NEW TAXONOMY OF (GRAY, SANTOS, & BIELOVA, 2023).

Although the taxonomy of Gray et al. is one of the most popular categorisations of dark patterns there are also other taxonomies that often add other dark patterns to their lists. Zagal, Björk & Lewis (2013) for example focus on dark patterns in game design. They mention *temporal dark patterns*, which are dark patterns that try to make players spend more time on the game, such as 'grinding', which means that a player has to repeatedly keep performing an action to make progress. Another category are *monetary dark patterns*, which try to make the player pay (more) for the game. A concrete form of this is the 'pay to skip' pattern, in which case the player pays an amount of money to get for example the opportunity to skip a level where they are stuck at. The last category are the *social capital-based dark patterns*, which are dark patterns that might put the players' social relationships at risk. This can for example be 'impersonation', in which the system connects names of real friends of the player to actions that they did not do themselves (Zagal, Björk, & Lewis, 2013).

The research of Mathur, et al. (2019) focusses on e-commerce websites. They add an extra classification to the existing taxonomies consisting of five dimensions, which are *asymmetric*, *covert*, *deceptive*, *hides information* and *restrictive*. 'Asymmetric' in this case means that choices offered to the user are provided in such a way that they are not shown equally compared to the alternatives, such as a opt-out button that is less visible than the opt-in button. A dark pattern is 'covert' when the design tries to steer a user into buying something without them recognising the effects of their choice, such as when an additional option is added to make others seem more appealing (decoy effect). 'Deceptive' is applicable when the user is confronted with misleading information, such as a countdown for a discount that is not really gone when the countdown has reached zero. When an interface 'hides information' it hides or delays relevant information from the user and an interface is 'restrictive' when it simply unnecessarily limits the choices a user has, such as only allowing specific types of signing up (Mathur, et al., 2019).

Bösch, Erb, Kargl, Kopp, & Pfattheicher (2016) present an overview of dark pattern strategies that are related to privacy. They mention known patterns such as 'privacy Zuckering' and 'bad defaults' (comparable to *preselection*), but also introduce five other privacy specific dark patterns. '*Forced registration*' is a pattern which means that a service is only available to users after registrations, which forces them to provide personal information. '*Hidden legalese stipulations*' are situations in which the terms and conditions are phrased in a way which makes it hard for users to understand what is in their, making them vulnerable for giving consent for something that they would not have consented to if they had known that is was in the terms and conditions. An '*immortal account*' is a dark pattern that refers to services that let their users never delete their accounts, or make this a very complex process, in order to try to prevent them from doing so. A variation on this is to make the user believe that all data is deleted, while the service actually keeps part of the data. '*Address Book Leeching*' is a dark pattern in which a user shares their address book with a service, for example to find friends that are also on the service, but at the same time the service is able to save all data (such as email addresses) of everyone in the address book. The last one of the privacy specific dark patterns mentioned are the '*Shadow User Profiles*' which is a phenomenon in which a service collects information about the users as well as the non-users. In a social media network the users can provide information about non-users, for example by sharing their address book, who can then based on the provided information be placed in the network, without them knowing or being informed about this (Bösch, Erb, Kargl, Kopp, & Pfattheicher, 2016).

In 2023 Gray, Santos & Bielova introduced a preliminary ontology of dark patterns (Gray, Santos, & Bielova, 2023). This ontology consists of the five categories of the taxonomy of Gray et al. (2018) complemented with the category social engineering. The ontology deviates on various points from the earlier published taxonomies. The *disguised ad* has earlier been labelled as interface interference and is in the new ontology categorised as sneaking. *Forced continuity* is categorised as forced action, but was earlier marked as sneaking as well.

2.1.4. Psychological background of dark patterns

As discussed in the previous section there is a wide variety of dark patterns that can be organised in different categorisations. There are also different mechanisms and theoretical backgrounds that explain the working of those patterns. This section will discuss various of the most relevant theories for some of the dark patterns.

Nudges. A nudge is defined as "*any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives*" (Thaler & Sunstein, 2008, p. 6). It hence focusses on small hints to put people into a certain type of behaviour and specifically does not focus on imposing limitations or restrictions on their behaviour. This type of techniques can be used within certain dark patterns, such as the *preselection* pattern, in which the users is steered towards a choice that is preferred by the owner of the system (Acquisti, et al., 2018), but no limitations are imposed as users are still free to change the *preselection*. This also applies to aesthetic manipulation patterns, in which for example one options if (un)favoured, for example having an unsubscribe button in small text and grey colour (Acquisti, et al., 2018). In this pattern users are again nudged towards a certain choice, but are not forced to do so.

Dual process theory. A well-known psychological theory is the dual-process theory. This theory states that humans use *type 1 processing* for quick and automatic reasoning, that requires little attention and use *type 2 processing* for more accurate reasoning, but which is also slower and needs more focused attention (Matlin & Farmer, 2017, p. 450; Kahneman, 2011). According to Bösch, Erb, Kargl, Kopp & Pfattheicher (2016) it is this theory that explains the working of some of the dark patterns. When creating an account for a website, one can use either type 1 processing, meaning that one does not read everything in detail and hence quickly and automatically accepts the terms and conditions, or one can use type 2 processing, meaning that one does read everything in detail, weighing the options, before either or not giving consent.

Type 1 processing is used especially when people have little motivation for something or have no ability to do something because they for example lack knowledge or time. This knowledge can be used by developers in the form of dark patterns. Someone needs to be already motivated to change the privacy settings, but might be overwhelmed by the complexity of it, lacking the ability to cope with this, and subsequently use type 1 processing instead of type 2, which might result in refraining from making any changes (Bösch, Erb, Kargl, Kopp, & Pfattheicher, 2016). Patterns related to this are *privacy Zuckering* and *hidden legalese stipulations*.

Cognitive biases. Humans often use 'heuristics' to make quick decisions. There is however the possibility that these heuristic result in wrong judgements, which is known as a cognitive bias (Maier & Harr, 2020).

One cognitive bias that is used in dark patterns design is the 'default effect', which is for example seen in the *sneak into the basket*-pattern (Mathur, et al., 2019) and *preselection*. For both these patterns it is the case that a choice is made for the user (either by placing an extra item in their shopping cart or by preselecting one of the options). The user has in both cases the possibility to change this (removing the item or changing the preselected choice), but the goal of the pattern is to steer the users into keeping the default extra item or the preselected choice. Research has shown that when there is one option selected by default, this increases the chance of people actually sticking with that choice and that "*opting-in does not equal opting-out*" (Johnson, Bellman, & Lohse, 2022, p. 13).

Another cognitive bias is the 'framing effect', which is for example seen in the *trick question*-pattern and in the patterns with aesthetic interference (Mathur, et al., 2019). These patterns try to steer users into making a certain choice or performing an action based on how something is presented. The wording of how a choice is presented matters in which decision people make. In an experiment by Tversky & Kahneman (1981) people were present with the casus: "*600 people are likely to die because of a new disease. There are two options: strategy A) 200 people will be saved or strategy B)1/3th probability that 600 people will be saved, 2/3th probability that zero people will be saved*". 72% of the people chose A. Another group was presented with the casus: "*600 people are likely to die because of a new disease. There are two options: strategy A) 400 people will die or strategy B)1/3th probability that no-one will die, 2/3th probability that 600 people will die*". In the second scenario only 22% of the participants chose A, although the outcome is exactly the same (200 saved, 400 die). This shows the strong influence of how choices are formulated on what choices people make in the end, even if the outcome would be the same (Tversky & Kahneman, 1981).

Low-stock messages, *High-demand messages* and the *Limited-time message* patterns are based on the 'scarcity bias' (Mathur, et al., 2019). In all those cases it is suggested that there is scarcity in the availability of a product, either in the form of a product being (in risk of getting) out of stock or its availability being limited in time. This should then give the user the sense of having to act fast in order to not miss out the product. The 'scarcity bias' is based on the psychological effect in humans that everything that is perceived as scarce automatically is seen as more valuable (Mittone & Savadori, 2009), which is artificially generated by suggesting that a product is scarce.

Social and emotional effects. There are various dark patterns with a social component in them, such as the social pyramid. Humans have a certain need to belong to others and if they are excluded this might influence their well-being. This also applies to humans in a digital environment (Williams, Cheung, & Choi, 2000). This fact is the foundation of some dark patterns, such as a form of *toying with emotion* by a social media website that stresses the number of friends that will miss you if you would unsubscribe from their service (Bösch, Erb, Kargl, Kopp, & Pfattheicher, 2016). Toying with emotion can also be applied by the way a text is stated, for example with statements like "accepting the cookies will enable extra functionalities", while ignoring negative consequences of accepting (Waldman, 2020). Another form of this is the dark pattern *confirmshaming*, in which the specific emotion of shame is used to steer users into a certain choice (Mathur, et al., 2019). An example of this is when an offer for a data protection programme can be declined by clicking a button with the text "No thanks, I do not want to protect my data" (Luguri & Strahilevitz, 2021, p. 62), in which the decision of the user to not start using the programme is framed in a way that it seems as if they do not care for their personal data.

Benefits and drawbacks in privacy choices. Some of the dark patterns have an impact on the privacy of the users, such as the forced registration. The functionality of this pattern is caused by the fact that the user is in a process of achieving a goal, in which the user is interrupted by the requirement of having to create an account. The user does so to achieve the goal, which would give 'instant gratification', which makes that critical thoughts on the privacy are ignored (Bösch, Erb, Kargl, Kopp, & Pfattheicher, 2016). People do in this case disclose personal data for the immediate benefit, whereas potential drawbacks of this dicsloure are typically experienced at a later moment (Waldman, 2020).

2.1.5. Occurrence and effectiveness

Various researches have been looking into the occurrence of dark patterns. A research from 2020 selected the 30 most trending apps from the Google Play store for each of the eight categories. It turned out that 95% of these apps contained at least one dark pattern, with an average of 7.4 (std. dev.: 5) dark patterns per app (Di Geronimo, Braz, Fregnan, Palomba, & Bacchelli, 2020). How often the different types of dark patterns occurred in this research is shown in Figure 3. It shows that *nagging*, *preselection* and *false hierarchy* are among the most used ones.

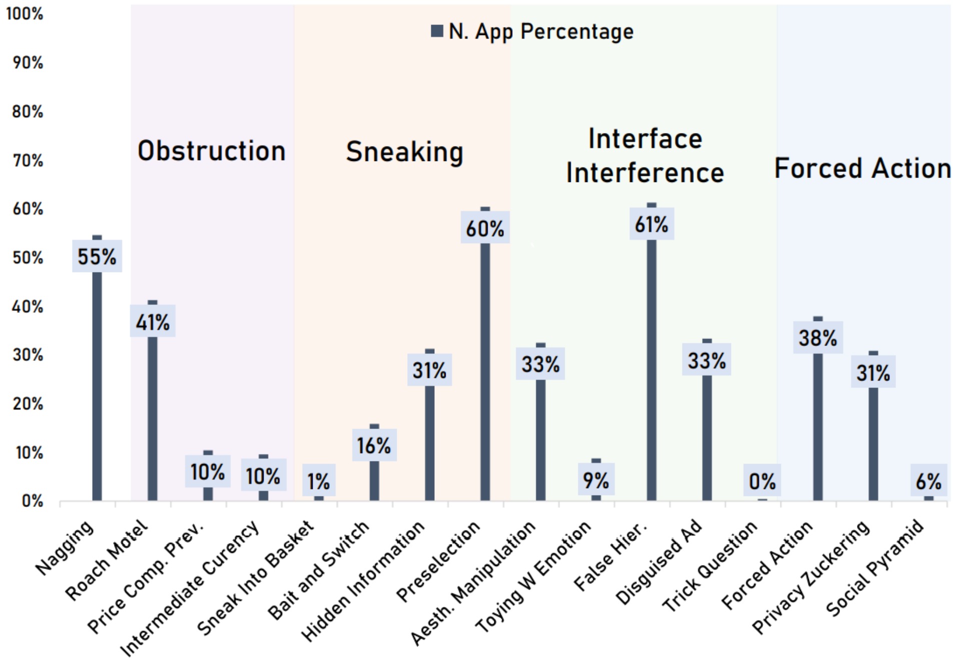

FIGURE 3 PERCENTAGES OF THE APPS CONTAINING A CERTAIN DARK PATTERN IN THE RESEARCH OF DI GERONIMO, BRAZ, FREGNAN, PALOMBA, & BACCHELLI (2020), IMAGE FROM (DI GERONIMO, BRAZ, FREGNAN, PALOMBA, & BACCHELLI, 2020).

Another research found 1818 dark patterns in a set of 11000 shopping website. This was based on an automated web search and hence only includes text-based dark patterns (Mathur, et al., 2019).

Sin, Harris, Nilsson and Beck (2022) showed that dark patterns can indeed be effective in increasing 'purchase impulsivity' compared to a control edition. This is confirmed by Luguri & Strahilevitz (2021), who showed that users in a dark patterns condition were – depending on the type of dark patterns used – twice to four times more likely to subscribe to a service compared to a control condition. Nouwens, Liccardi, Veale, Karger, & Kagal did an experiment in which they removed in the experimental condition the opt-out button on a privacy consent form from the first page, which resulted in an increase of about 22 percentage points (2020).

2.1.6. Legislation

There are legal regulation that determine what is allowed on the web. In the European Union for example the GDPR applies, in which is stated that processing of personal data is only allowed when the one who's data is processed has given consent to do this, or any of the other requirements for legal processing has been met (such as processing based on a contract or because of a legal obligation, mentioned in article 6) (GDPR, 2016). This means that in general permission of the user is required when using techniques likes cookies and trackers. To get the permission of the user consent management platforms (CMP) have been introduced, in which users should be able to freely give unambiguous consent (Nouwens, Liccardi, Veale, Karger, & Kagal, 2020). The use of prechecked checkboxes in a consent form is in the GDPR (recital 32) explicitly mentioned as no valid freely given consent. Moreover should it according to article 7 also be possible to withdraw consent after it is given, in which withdrawing consent should be as easy as giving it (GDPR, 2016).

Research by Nouwens et al. (2020) used scraping techniques to analyse implementations of five different CMPs on popular websites in the United Kingdom (N = 680). Their research has shown that only 11.8% of the sample met the minimum requirements of the GDPR. Part of the problems with not meeting the requirements can be related to dark patterns. 56.2% of the websites had a consent form containing preticked options, which can be related to the dark pattern *preselection*. Moreover it is suggested that consent walls that only allow access to a website if consent is given are in some situations not a valid way to ask consent to website visitors (Zuiderveen Borgesius, Kruikemeier, Boerman, & Helberger, 2017). This is a form of *forced action*.

An experiment by Nouwens et al. (2020) showed that if the button 'reject all' is not shown on the first page of a cookie consent form the chance that a user provides consent is higher. It can be argued that this is a violation to the degree of which the consent is freely given.

Moreover the new Digital Services Act of the European Union introduces new rules with respect to dark patterns (Gunawan, Santos, & Kamara, 2022). Article 25 of the law for example states that providers of online platforms are not allowed to use interactions that deceive or manipulate the users (DSA, 2022). Also recital 67 of that same law states that service providers should not present choices in a non-neutral way using visual components, if this is not in the interest of the user (DSA, 2022), which is what the dark pattern *false hierarchy* does.

2.2. Performing tasks in a digital environment

The second part of this literature section discusses how various elements affect the time it takes a user to perform an action within a digital environment. This is relevant background information for RQ1.

2.2.1. Process of a task

When a person wants to perform an action, one goes through various stages of executing and subsequently evaluating that action. Donald Norman has developed the Seven Stages of the Action Cycle to describe this sequence, which is graphically shown in Figure 4. On the execution side ("bridge of execution") there are four stages, which are the *goal* that someone has in mind, which is followed by a *plan*, which one then *specifies*, and finally *performs*. This action has a certain impact on the world (for example someone turns a light on or clicks a button on a website), after which one evaluates what happens ("bridge of evaluation"). The evaluation side consists of three stages, which are *perceiving* what the new state of the world is, *interpreting* this new state and finally *comparing* the outcome with the original goal one had in mind (Norman, 2013).

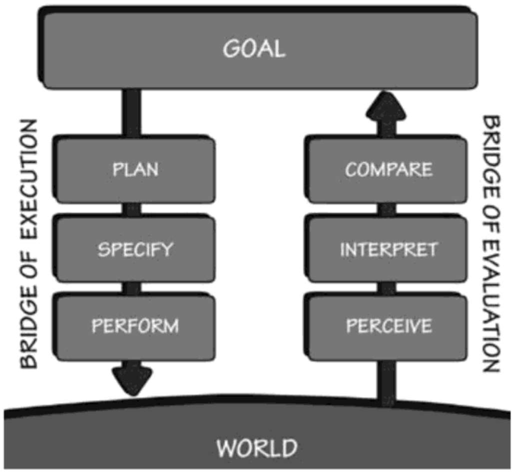

FIGURE 4 SEVEN STAGES OF THE ACTION CYCLE, IMAGE FROM (NORMAN, 2013).

A comparable but still slightly different model of how an action works is the theory on 'microinteractions', which are the "*the functional, interactive details of a product*" (Saffer, 2014, p. 3). This model is hence also more specific than the seven stages model. A microinteraction starts with a *trigger*, which can be user-initiated (the users wants achieve a goal and starts doing something) or system initiated (a new message arrives and the system shows a pop-up). The next step consists of *rules*, which determines which behaviour occurs after a certain action. The following part of the process consists of *feedback*, which is anything that lets the user know what has happened. This can be visual, such as a message on a screen, but also aural or haptic. The fourth and last step of a microinteraction are the *loops and modes*, which respectively explains what happens after the interaction is finished (does it for example expire after some time?) and whether there are situations in which the interaction should behave in a different way than it normally would (Saffer, 2014).

For the time spent on a webpage one can focus on a larger scale (the time an individual spends on a single page, possibly executing multiple actions), or a smaller scale (more zooming in to a single action, such a clicking or typing). Regarding the first one, research has been done about the time spent on webpages (known as 'TSP') (Nagy & Gaspar-Papanek, 2009). One of the factors that has an influence on the time spent on webpages is the type of page (such as an informational page, a contact page or a product page), but also the quality of the page (layout and design) and naturally the length of the page (Hofgesang, 2006). This is confirmed by research by Choi, Seo & Lee (2009). Their research also shows that interest of the user and credibility level of the webpage are important contextual factors that influence the viewing time. Next to that the task and the language also affect viewing time. The research could not show that difficulty or complexity levels have an effect on the viewing time. Hofgesang (2006) also noted that the reading speed of the user and the speed of the server and network can affect the measured time spent on the webpage in researches.

Bhatnatar, Sinha & Sen (2019) proposed a model in which they state that visit duration is dependent on website trust, site attractiveness, information quality and personalisation. Their research adds to that that the navigational ability – which itself is dependent on the easy of use – also effects the visit duration.

Focussing on the smaller scale, not much is yet known about how much time single actions or the stages of these specific small actions cost in a digital environment. Lam (2008) has proposed a model that maps the stages of Norman's model (2013) to stages that all have certain *interaction costs*, but these are not directly related to units of time. Norman (2013) however notes that a factor as experience can influence how an individual goes through the stages of earlier mentioned seven stages of the action model. An experienced driver for example can almost automatically turn right, whereas someone who is learning to drive has to think about all the steps to be taken (Norman, 2013).

2.2.2. Visual gaze patterns in a digital environment

Another part of the processing of a task is the perceiving of the visual information that is shown. When people read a text their eyes move with a series of jumps over the line, which is known as *saccadic eye movement*. Between those movements there are small breaks, known as *fixations*, in which information is processed by the visual system (Matlin & Farmer, 2017, p. 88).

When people perceive a website they often first read the top of the page from left to right, then move a bit lower and read again from left to right and then move down over the page, forming an F with their eye pattern. This is hence known as the *F-Shaped Pattern*. This pattern also shows that users do not read a text on a website always thoroughly (Nielsen, 2006). Figure 5 shows examples of the F-Shaped Pattern, based on a heatmap from an eye tracking research.

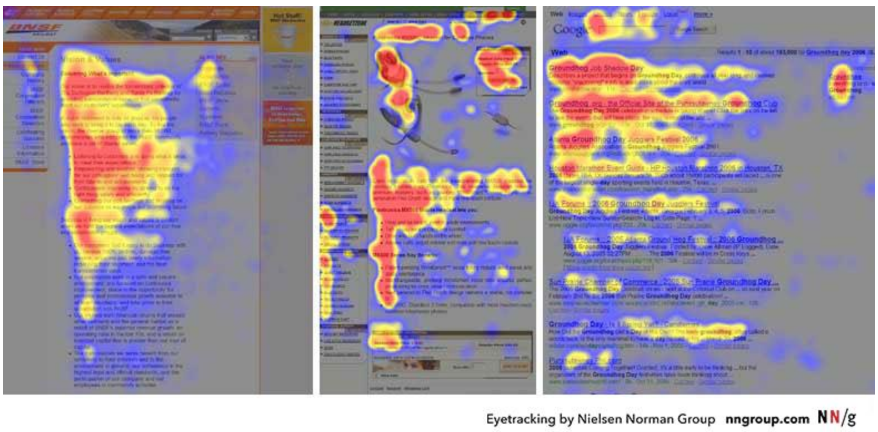

FIGURE 5 EXAMPLES OF THE F-SHAPED PATTERN, IMAGE FROM (NIELSEN, 2006).

There are also other known visual gaze patterns that occur on webpages. One of these is the *spotted pattern*, which means that the user fixates on a region on the page either because it stands out (for example because it is a link or it is coloured) or because it looks like something they are looking for (such as a number when they are searching a phone number). Another pattern is the *layer-cake scanning pattern* in which a pattern is seen of the user who is mainly reading headings, skipping the text and reading the next header, until they have found something they are interested in. The *commitment pattern* is seen when a users reads the page, instead of merely scanning the content. This is seen as a visual gaze patterns with fixations on (almost) every word (Pernice, 2019). Another pattern that is seen on pages with a lot of text, or other content that is evenly displayed over the page and is of comparable importance to the user is the *Gutenberg pattern*. This pattern describes a movement from the top left at the beginning to the lower right part of the page (Hernandez & Resnick, 2013).

The *Golden Triangle pattern* is seen when a user starts in the left top of the page, then moves to the right and finally moves (diagonally) down back to the left of the page (Hernandez & Resnick, 2013).

How people scan or read a web page depends on a variety of factors, including the task or goal they have, their assumptions and previous experiences, the layout and the content of the page (Pernice, 2019). Next to that it is also important to note that not everyone uses a webpage in the same way and that there are differences among people in gaze patterns and in the parts someone pays the most visual attention to (Dumais, Buscher, & Cutrell, 2010). One factor that can have an effect on how people use a webpage is age, as eye tracking research has shown that different generations had different viewing behaviour (Djamasbi, Siegel, Skorinko, & Tullis, 2011).

Although not very much is known about the specific visual gaze patterns for various interactive elements in interfaces, there are some general characteristic known about the way people interact with certain elements. An older study by Goldberg et al. (2002) focussing on a webpage with 'portlets' (user-customisable boxes) showed that users had a slight preference to searching across different columns (horizontally) instead of searching within the same column (vertically). Eye tracking research has also shown that when group of buttons in an interface are well-organised this requires shorter scanpaths and less fixations compared to the buttons being randomly placed (Goldberg & Kotval, 1999).

2.3. Introduction to the research

As mentioned in section 2.1.2. there has been an increase of research on the topic of dark patterns. Various research lines in this area were also mentioned at the workshop *Future Proof Methods for Measuring and Detecting Dark Patterns* [9]. This section will discuss the research goals of the current research, preceded by a discussion of the distinction that will be made between timebound and local dark patterns.

2.3.1. Timebound and local dark patterns

As this research is about the effect on time needed for a task and the effect on visual gaze patterns in situations in which dark patterns are applied, it should be noted that not all types of dark patterns are applicable for researching this topic, as they are more focussed on an effect on the longer term. These patterns are less of interest for the type of research proposed in this document, as it would be hard to draw conclusions about the time they cost a user, given that this can not be measured at a single moment.

[9] 19th of October, 2022 in Utrecht, The Netherlands.

The two types of dark patterns are in this research defined as *timebound dark patterns* and *local dark patterns*. How the dark patterns can be divided among those two categories is shown in Figure 6.

Timebound dark patterns		**Local dark patterns**	
Roach motel	Intermediate Currency	Nagging	Price Comparison Prevention
Forced Continuity	Gamification	Hidden Costs	Bait and Switch
		Hidden Information	Preselection
		Privacy Zuckering	Forced Registration
		Countdown Timer	Limited-Time Message
		Low-Stock/High-Demand message	

FIGURE 6 CATEGORISATION OF TIMEBOUND AND LOCAL DARK PATTERNS.

Timebound dark patterns. Timebound dark patterns are the types of dark patterns that involve a certain period of time between their initiations and their (full) effect and typically exist of multiple actions. An example is the *roach motel* pattern. This dark pattern makes it very easy for someone to subscribe to a service, but when someone wants to unsubscribe after a few months it turns out that it is very hard to do so. In this case a longer period of time is included (the time between subscribing and unsubscribing) and it consists of multiple actions (in this case subscribing and unsubscribing).

A dark patterns that can also be placed in this category is *intermediate currency*. The goal of this pattern is to make users buy virtual money, which they value differently compared to regular money. This might cause the user ending up with a different way of spending. Also for this pattern it holds that the effect takes place over some time and that multiple actions are involved (buying the money and at a later moment spending (more of) it).

Local dark patterns. Local dark patterns are the types of dark patterns that only present themselves in one moment and have their effect at that same moment or directly on or after the interaction that follows on the presented dark pattern. An example is for example *preselection*, in which one of the options already is selected by default in the hope that the user will stick with that choice. This pattern is local as it only presents itself at the moment the choice is made. The *preselection* only affects the user from the moment the choices are presented and does no longer effect the user after the choice is made. The choice itself of course can have consequences at a later moment, but the dark pattern itself only is influencing the user for a short period.

Another pattern that can be seen as a local dark pattern is the *bait and switch* pattern. This pattern causes elements to perform another action than the action the users had expected. This dark pattern again consists of one single action and happens on a single moment.

Some patterns are related to privacy, such as *privacy Zuckering*, *hidden legalese stipulations* and *forced registration*. These type of patterns are all focussed on getting more personal information from the users than they were intended to provide. One could argue that these patterns should be labelled as a timebound dark pattern, given that the privacy violations takes place over a longer period of time. However from a user-centred perspective it is more logical to see these as a local dark pattern, given the fact that the user is involved in a single action on a single moment (for example by giving consent against their intention or registering for a service while they would rather not do so). This follows the same line of reasoning that says that *preselection* belongs in this category, regardless of what effect the choice made can have in the future, as there is only one moment in which the user interacts with the system, being confronted with a dark pattern.

2.3.2. Research goals

The goal of this research is to get more insight in the amount of time a users needs to cope with a dark pattern and to show how they affect the visual gaze patterns of a user. This is based on RQ1 (h*ow do dark patterns affect the time users need to complete a task?*) and RQ2 (*what visual gaze patterns can be seen around applications of dark patterns?*) as presented earlier.

RQ1. Section 2.2.1. has shown that there are various factors that influence how much time someone spends on a webpage. Among these factors are the type of page, trust and information quality. Some of these factors might be influenced by the use of dark patterns. Research by Maier & Harr (2020) showed that if a company uses too many dark patterns the trust that people have in it might be decreased. For some of the dark patterns mentioned in section 2.1.3. it can also be argued that they affect the information quality, as bad quality of information is part of their workings. This for example applies to *hidden costs* (part of the category 'sneaking') and *hidden information* (part of the category 'interface interference'). As these mentioned factors change based on the presence of dark patterns, this would according to the aforementioned research also affect the time a user spends on a page.

When focussing on the psychological side of the dark patterns, an effect on the time an action takes can be explained by looking at psychological models. Effects on the time needed can be explained by assuming that a psychological process underlying an action – such as the processes described by Norman (2013) and Saffer (2014) in section 2.2.1. – is interrupted when interactive content behaves in a different way than expected, for example if the user is confronted with dark patterns as *nagging* or *bait and switch* (part of the category 'sneaking').

This leads to the following hypothesis for RQ1:

H0: The use of dark patterns has no effect on the time a user needs for a task compared to the same situation without dark patterns.

H1: The use of dark patterns does have an effect on the time a user needs for a task compared to the same situation without dark patterns.

RQ2. Section 2.2. has discussed how a digital task is processed by humans, with a specific focus on how the visual gaze pattern in this functions. As far as known there is no specific research on how these are seen around applications of dark patterns.

To investigate these visual gaze patterns this research will include an eye tracking study on applications of dark patterns to understand how they are perceived by users and how this effects their viewing. Section 2.2.2. has discussed various visual gaze pattern, such as the F-shaped pattern or the golden triangle pattern. Based on the data of the eye tracking study heatmaps for the same experimental website with and without dark patterns can be generated. These can provide insight in whether some of the known patterns are seen and whether there are differences between the two versions (i.e. whether dark patterns have an effect on these viewing patterns).

Next to qualitatively analysing the patterns seen, there are various metrics that can be compared to provide additional insight in the way dark pattern influence users' viewing. These metrics involve for example the time the eyes of a user fixate in a certain Area of Interest (AOI-time) or the time it takes before a user fixates in a relevant area (such as an aera with a dark pattern) for the first time (TTFF). The setup will be discussed in detail in section 3.

3. Methodology

This section describes the way the practical part of this research was caried out in order to answer the research questions proposed in section 1.

3.1. Research design

The practical part of the research consisted of an experiment in which the effects of dark patterns were tested. To be able to draw reliable conclusions about the effects of dark patterns this had to be compared to a similar situation without dark patterns. Hence the experiment had two conditions: one with and one without applications of dark patterns. The experiment design was a combination of *within-subjects* and *between-subjects* as explained in the next paragraph.

As it is likely that it would have affected the results if the participants had performed the same task twice, there were two tasks with both a condition with and a condition without dark patterns in the design. This means that there were four designs in total (2 tasks * 2 conditions) of which each participant saw two designs. Each participant saw one task in one condition and the other task in the other condition (so if task 1 had no dark patterns, task 2 would). The order of the task was randomised, so some participants started with task 1 and others with task 2.
The randomisation of the research is shown in Figure 7. Each participant will be randomly assigned to one of the four flows shown.

Task 1 (DP) → Task 2 (C)	Task 1 (C) → Task 2 (DP)
Task 2 (DP) → Task 1 (C)	Task 2 (C) → Task 1 (DP)

Figure 7 Task and condition randomisation options.

The participants were asked to perform the given tasks in web environments designed for this experiment. Their process was screen-recorded and the participants were eye-tracked. Afterwards the recording of the dark pattern condition were watched together with the participant, during which they were asked to elaborate on their actions. The process of the experiment is described in more detail in section 3.3. The material (both hardware and software) used in the experiment is discussed in more detail in section 3.4.

3.1.1. Eligible dark patterns

As discussed in section 2.3.1. a distinction can be made between local and timebound dark patterns. Only local dark patterns are eligible to be part of the experiment, as it will not be possible to measure an effect over a longer time (e.g. the roach motel only works if people subscribe to something and for example a few months later want to unsubscribe).

There are also other dark patterns that are less convenient to use in the experiment. The *disguised ad* pattern leads users to an (external) advertised website when they click on it, which is not desirable in the experiment, as the participants will then fall out of the flow of the experiment. Also *hidden costs* might not work well in the experimental setting, as users do not have to pay actual money, so they will probably react differently then they would have if they actually had to pay the hidden costs. The same holds for *forced registration* and *privacy Zuckering*, as in the experimental setting participant will know that their data is only used within the experiment.

3.1.2. Tasks used

The two scenarios used in the experiment will be described in this section, together with the dark patterns that were used in the dark patterns condition.

Task 1. In the first task the participants were asked to book a train ticket for a journey from Rotterdam to London. The design of the website used was inspired on the website of international ticketing service of the Dutch Railways, NS International (https://www.nsinternational.nl). This task was chosen as users of such a service have to make various choices during the process that can be influenced by dark patterns. Table 2 gives an overview of the steps of the task and the dark patterns that were included in each step in the experimental condition.

Task: "Book a train journey from *Rotterdam Central* to *London St. Pancras International* on the 15the of March, 2023, with departure time 14:28. You travel alone (1 person) and want to travel 2^{nd} class. You want to book the train journey *only* and do *not* need any extra (such as a cancellation insurance). (The task ends at the payment)".		
Step	**Action**	**Dark pattern**
1	Enter origin and destination	-
2	Select date from a calendar	-
3	Select journey/time	-
4	Select travel class	False hierarchy
5	Reserve a seat (optional, extra costs)	Sneak into basket
6	Decide on cancellation insurance	Trick question
7	*Task completed*	

TABLE 2 FLOW OF TASK 1 WITH DARK PATTERNS APPLIED IN IT.

Upon starting the first task, the participant saw the home screen of the mock-up train ticket selling service, called 'TrainDiscounter.com', as shown in Figure 8. Here the participant had to enter the given origin and destination and click 'search'.

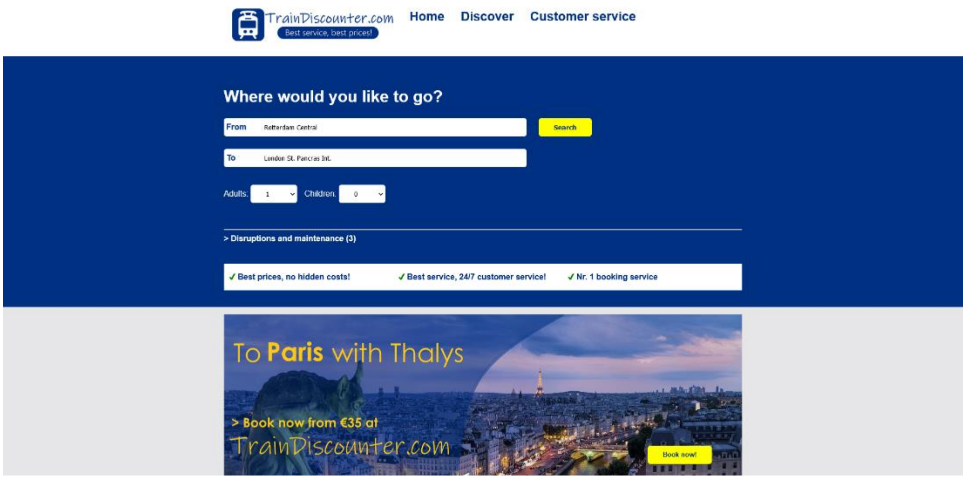

FIGURE 8 FIRST SCREEN OF TASK 1.

After selecting the given date in the next screen, the participant had to select a journey. This led the participant to an checkout screen in which various choices could be made, as shown in Figure 9.

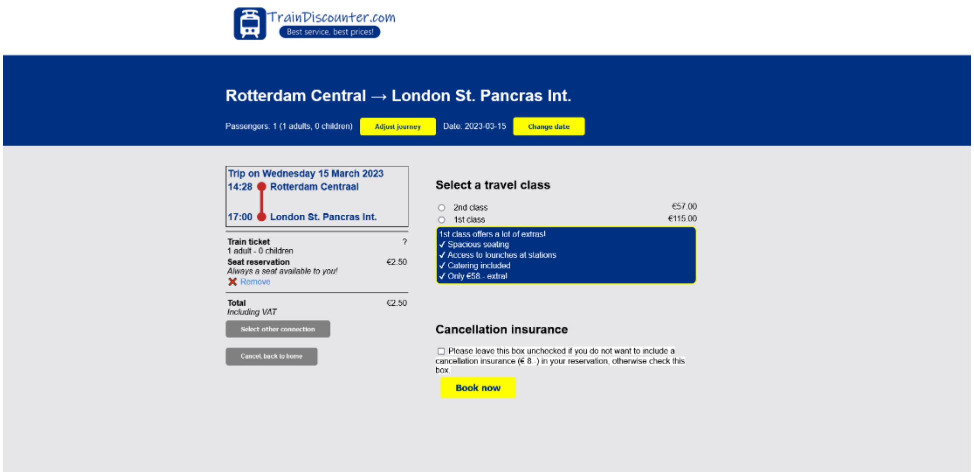

FIGURE 9 CHECKOUT SCREEN IN TASK 1.

The checkout screen showed the selected journey and offered the user various choices. In the experimental condition the design of this page included three instances of dark patterns.

First a *false hierarchy* that tried to promote first class over second class. This was done by placing a striking box with advantages of the first class with this option, in order to make it more attractive than the second class option. In the purchase overview on the left of the screen an instance of *sneak into basket* tried to include a seat reservation without the user knowing. Finally there was a *trick questions* that asked in a convoluted way ("*Please leave this box unchecked if you do not want to include a cancellation insurance (€ 8.-) in your reservation, otherwise check this box.*") whether the user needed an cancellation insurance (which they did not based on the instructions).

The task ended when the participant clicked 'book now' on this screen; the payment and completion of the reservation were not part of the task.

In the control version the majority of the task flow stayed the same. The steps up to the checkout screen were exactly similar. In the checkout screen all dark patterns were removed: the selecting of the travel class did not favour one option of another, no seat reservation was made automatically (although an option to add one was visible) and the checkbox to include a cancellation insurance had a much simpler phrased label ("*Include a cancellation insurance (€ 8.-) in your reservation.*"). This task also ended once the participant clicked 'book now'.

Task 2. In the second task participants were asked to book a hotel for two nights in Berlin. The design of the website used was inspired on the website of Booking.com (https://www.booking.com). The task was chosen as it also required the participants to make some choices during the process, that could be influenced by dark patterns. The steps and dark patterns used in the experimental condition were as shown in Table 3.

Task: "Book a hotel in Berlin; you want to check in on the 17th of March, 2023 and check out on the 19th of March, 2023 (=2 nights). You travel alone (1 person). Choose the hotel the closest to the city centre. You want to book *one standard room*. (The task ends as the payment)".		
Step	**Action**	**Dark pattern**
1	Cookie pop-up	Preselection
2	Choose destination and date	-
3	Pop-up with advertisement of other trips	Nagging*
4	Select the hotel	High-demand/Low-stock messages
5	Hotel page, selecting a room	Toying with emotion
7	Payment and completion	-

TABLE 3 FLOW OF TASK 2 WITH DARK PATTERNS APPLIED IN IT. * AS THIS STEP ITSELF IS A DARK PATTERN IT IS LEFT OUT IN THE CONDITION WITHOUT DARK PATTERNS.

When the participant started this task, first a cookie pop-up was shown. This pop-up allowed a user to choose between accepting and declining cookies, and to confirm their choice with a button. As a dark pattern *preselection* was used, which means that the accepting option was selected by default. This is shown in Figure 10.

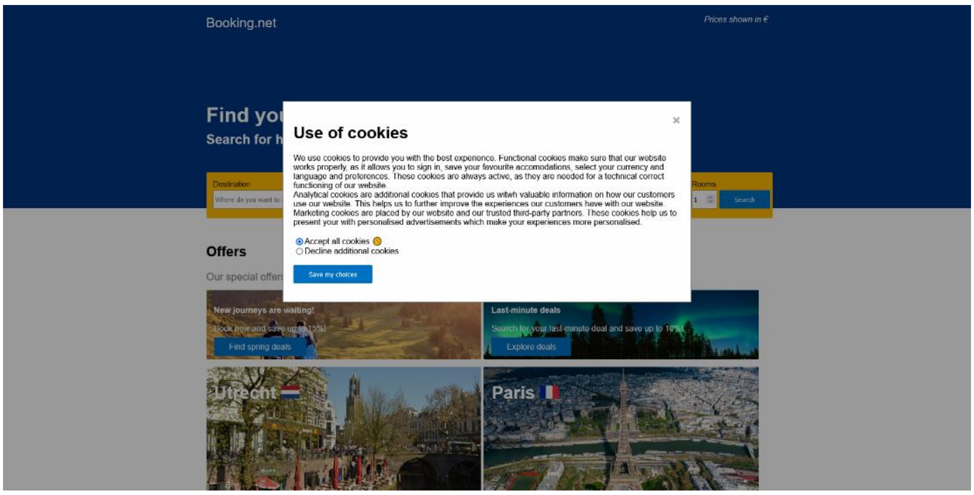

FIGURE 10 COOKIE POP-UP IN TASK 2.

In the next step the participants had to enter the destination, dates and number of travellers and click ‘search’. Before the results were shown a pop-up with an advertisement is shown, as a form of *nagging*. This is shown in Figure 11.

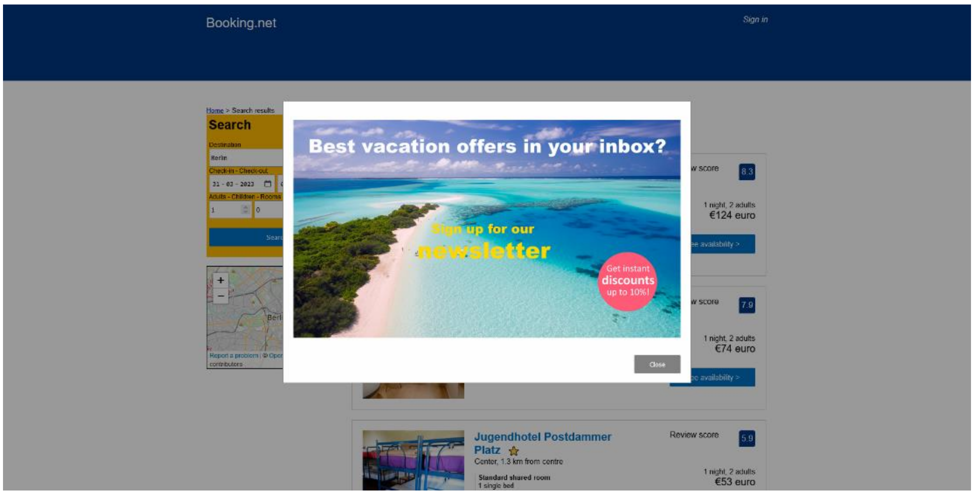

FIGURE 11 *NAGGING* IN TASK 2.

The participants were asked to select the hotel that is located the closest to the city centre, which they could do based on the distances in kilometres from the centre shown for each hotel. In the information boxes for the hotel *high-demand* and *low-stock* messages were displayed for some of the hotels (e.g. “only 7 rooms left at this price on our site!” or “booked 4 times in the last 24 hours!”). This is shown in Figure 12.

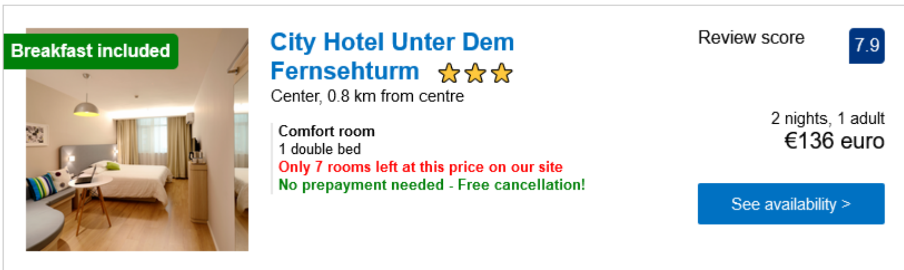

FIGURE 12 *HIGH-DEMAND AND LOW-STOCK MESSAGES*.

Once a hotel was selected the participant arrived at the ‘hotel page’, on which information and pictures of the selected hotel were shown. On this page the booking could be confirmed by selecting the desired type of room. Here the *toying with emotion pattern* (the specific way it was used here is also known as *confirmshaming*) tried to persuade the participant to select the more luxurious rooms. Once the participant moved their mouse cursor over the reservation area a nudge appeared that tried to persuade the user to choose a luxurious room. To close the nudge the participant had to click on a link saying “No, thanks, I would rather skip this offer and have a simpler experience”, as shown in Figure 13.

FIGURE 13 TOYING WITH EMOTION TO PERSUADE USERS TO CHOOSE A MORE LUXURIOUS ROOM.

In the control version the basic process of the task stayed the same. The cookie pop-up was still part of the task, but did not have a preselected option. The advertising pop-up (*nagging*) was left out of the task. A drawback of this is that it does not offer the best options for comparison between the two versions, but as the appearance of the pop-up *is* the dark pattern in this case, there is no other option to compare it with a version without this pop-up at all. The various hotel options were shown in the same way in the control condition as in the experimental condition, but none of the hotels had *high-demand or low-stock messages* in the control condition. Finally, the participants were also in the control condition being persuaded to choose a luxurious room, but the closing link was changed into a simple 'close'-link, in contrast to the experimental condition in which the link was trying to address the participant's feelings.

3.1.3. Retrospective think-aloud

After the participant had completed the tasks, a brief retrospective think-aloud session followed. In this session together with the participant the video of the task in which dark patterns were applied was being watched. On this video the eye movement was also visible. While watching the recording the participants were asked to explain what they were doing, thinking, feeling, etc. at the moments shown in the video. This setup has in earlier research proven to have participants give more honest and in-depth answers when they see their own eye tracking recording compared to only asking them (Cho, et al., 2019, p. 369). The sessions could be either in Dutch or English, based on the preferences of the participant.

During the retrospective think-aloud session the audio was recorded, if the participant had given consent to do so. If a participant preferred to not have their voice recorded, handwritten notes would be taken. For the sake of time needed for the whole experiment, only the task in which dark patterns were implemented were watched. As the interest of the research mostly lies in how people perceive dark patterns there is less need to also watch the control version.

3.2. Participants

Participants were recruited based on convenience sampling. There were no specific requirements a participant had to meet before participating, however participants needed to be at least 18 years old and have no severe sight problems. Participants who are younger than 18 years old had to be excluded as this would have required additional ethical measures and (possible) permission from parents or guardians.

Participants in the experiment needed to have normal or corrected to normal eye-sight. Severe visual problems might cause problems with the eye tracking devices. Glasses or contact lenses however do not cause problems.

Convenience sampling was used to gather participants. Participants were not informed on the actual goal of the study (i.e. investigating dark patterns) but were invited to take part in a more general described study ("an eye tracking study to research how people interact with websites"). It is however possible that participants might have recognised the actual goal of the study if they were familiar with the topic of dark patterns.

Before the actual experiment took place two pilot studies were done. The first one without using the eye tracking device to test the material and the second one to test whether the eye tracking device functioned as expected. The data of the people participating in the pilot study was left out of the actual data set.

3.3. Experiment process

This section will describe the experiment process. The experiment took place in the Human-centred Computing Lab in the Buys Ballot building in Utrecht.

3.3.1. Before the experiment

- Participants were invited to participate. They received a link with which they could schedule their participation within the two weeks the experiment ran.
- Participants were with the invitation already informed about the general idea of the experiment (except for the exact goal) and are told about:
    - The requirements they had to meet to participate (at least 18 years old and no severe eye problems);
    - That they would perform a task while being eye-tracked and that they would be asked some questions afterwards;
    - The location of the experiment.

3.3.2. During the experiment

- The participant entered the Human-centred Computing Lab of Utrecht University and was welcomed and thanked for their participation. Participants were informed about what would happen during the experiment.
- The participant received the information sheet and consent form and got some time to read it, ask questions and when they agreed to participation to sign it.
- First the participant was asked to complete a short survey with demographic questions. Those questions involved gender, age and educational level.
- Subsequently the eye tracking software was started. The participant was asked to sit in a position that was comfortable, after which the distance and angle of the screen were changed to make sure the eye tracking device could see the participant's eyes.
- After this setup was completed, the participant was asked to read the instructions of the first task from the paper. After these had been read, the participant could click on the 'start' button themselves, after which the calibration started and once finished the first task could be executed.
- Once the participant had finished the first task, there was a short break.

- During the break the participant was also asked to read the instructions for the second task, which was also being executed by the participant once they finished the break.
- Together with the participant the eye tracking recording of the task which involved dark patterns was watched. The participant was asked for commentary on what they saw, did, and thought. If the participant agreed on using voice recording while signing the consent form, this was also being recorded, otherwise handwritten notes would have been taken.

3.3.3. After the experiment

- Participants were thanked once more for their participation.
- Participants leave the lab after being offered a bar of chocolate as a thank-you.
- The data of the session (both the eye tracking data as well as the voice recording of the retrospective think-aloud session) were uploaded to a cloud service of the Utrecht University and deleted from the recorder and local folders of the computer.

3.4. Setup and material used

This section will discuss what the experiment looked like from a practical perspective. It describes both the hardware and software used to make the experiment possible.

3.4.1. Hardware and location

The experiment took place in the Human-Centred Computing Lab of Utrecht University in the Caroline Bleeker Building in Utrecht. For this experiment a Windows 10 desktop computer (64 bit) was used, to which a screen-based eye tracker was connected. This device was a Smart Eye AI-X. The monitor resolution was 1920 * 1080, 60Hz.

The retrospective think-aloud session was recorded, if the participant gave permission for this. If this was the case the recordings were made using a mobile phone and immediately after the session the data was saved to a university computer.

3.4.2. Software

The software that was used for capturing, storing and processing the eye tracking data was iMotions.

For the task itself custom webpages were created, using HTML, PHP, CSS and JavaScript. These pages were developed with and without dark patterns. The pages were stored on a Utrecht University server.

3.5. Data and analytics

This section will discuss the variables of the data collected and how the data was analysed.

3.5.1 Variables

The independent variable in the experiment is the condition in which a participant is. This variable is either ‘with dark patterns’ or ‘without dark patterns’. The other independent variable is the type of dark pattern that is used. The types used are *sneak into basket*, *trick question*, *false hierarchy*, *preselection*, *nagging*, *low-stock/high-demand messages*, and *confirmshaming*. There are various dependent variables in the experiment, that are either relevant for RQ1 or RQ2. These are discussed in this section.

RQ1. First of all the *time* was measured. This variable was measured in (milli)seconds. With the help of the eye tracking software both the total duration of a task as well as parts of it (called scenes) were measured. Especially the scenes where participants interacted with dark patterns were of interest. For each dark pattern the time of the scene in which it occured was used as the measured 'time spent'. This time would be compared against the control condition.

Nex to the time the gaze data of the participant was also stored by the eye tracking software. It recorded gaze points, which were all the locations of the eye of the participant on the screen. This was stored 60 times per second automatically by the eye tracker. A series of gaze points (typically 100 to 300 milliseconds) on the same location is known as a fixation. Movements of the eye between those fixations are saccades (iMotions, n.d.).

The previously mentioned data points for tracing eye movement were mainly raw data points, but there were also more advanced metrics that could be calculated based on these. One way to do this is by using predefined *Areas of Interest (AOI)*. An AOI can for example be drawn around a menu, image or button. With this data for example the time spent looking within a specific AOI could be analysed (iMotions, n.d.).

RQ2. The *visual gaze patterns* are another way of looking at the results of the experiment. Here the main independent variables were also whether someone is in the control or experimental condition, and the type of dark pattern that is applicable. The dependent variable is the way the eyes of the participants moved over the screen. During the analysis the patterns that occur in those movements would be categorised.

3.5.2 Analytics

After the experiment had been finished the data was analysed. The general way of analysing is explained in this subsection. The next chapter (chapter 4) describes the exact execution of the analysis.

RQ1. To answer RQ1 the time spent in the control condition had to be compared with the times in the dark patterns condition. For each of the two tasks separately the average amount of time needed by the participants had to be calculated. The time difference between the two tasks would be statistically tested by using an Independent Samples t-test (or a Mann-Whitney U-test if the values required a non-parametric test).

For the individual dark patterns a comparable analytical strategy would be applied. Instead of the total time of a task each of the scenes involving a dark pattern would be analysed separately.

Next to that, AOIs were drawn around areas in which dark patterns were applied. In the control condition they were drawn around the element that included a dark pattern in the experimental condition. A selection of relevant metrics of the AOI were compared between the two condition. To do this an Independent Samples t-tests (or Mann-Whitney U-tests as non-parametric alternative) was be used.

RQ2. For the gaze patterns qualitative analysis was applied. For each of the dark patterns gaze patterns that were (repetitively) seen were labelled and named. Afterwards the number of occurrences for each type of gaze patterns were counted.

3.6. Ethics

The planning of the experiment involved some ethical considerations. Normally it is for example the right of a participant to exactly know what the study is about. In this experiment this was however impossible, as otherwise chances were high that this would have influenced the behaviour of the participants (e.g. ignoring or paying extra attention to the dark patterns). Hence the participants were not told about the exact goal of the study on beforehand, but were informed about this afterwards.

To guarantee the privacy of the participants no data with which they could directly be identified was collected as part of the experiment. The data needed for planning the experiment sessions (appointment data and email address) were destroyed within two weeks after the experiment and not connected to the session of the participant. The audio recording of the retrospective think-aloud session was optional and for the participants that give permission for audio recording, the recording was transcribed and deleted within two weeks after the experiment.

The experiment plan was submitted to the Ethics and Privacy Quick Scan of the Research Institute of Information and Computing Science of Utrecht. The moderator of the Human Computer Interaction programme gave permission for the research to be started.

4. Results

This section will discuss the results of the experiment. First the demographics of the sample are discussed, after which the results of the experiment are reported. In the subsections first time based results are discussed, then the results with respect to the Areas of Interest and finally the results regarding the gaze patterns are reported.

All reported numbers in this section are rounded to (max.) three decimals.

4.1. Demographics and conditions

The experiment was executed between the 28th of February, 2023 and the 10th of March, 2023. In total 13 people participated in the research. The majority of the participants were female (9), a smaller part were males (3) and one non-binary person participated. The most participants were in the age group between 21 and 30 years old. Table 4 gives an overview of the age distribution of the participants.

Age group	**18-20**	**21-30**	**31-40**	**51-60**	**60+**	**Total**
Number of participants	1	8	2	2	0	0

TABLE 4 OVERVIEW OF THE AGE GROUPS OF THE PARTICIPANTS.

Almost all participants (11 out of 13) had completed or were currently enrolled in a university master programme. The two other participants had completed or were currently enrolled in a university bachelor programme and a higher professional education programme.

7 participants performed the first task in the control condition and the second task in the experimental condition. The other 6 participants performed the tasks the other way round: the first task in the experimental condition and the second one in the control condition.

4.2. Data processing and preparation

This subsection describes how the data from the experiment was processed and prepared for analysis.

4.2.1. Execution of the experiment and abnormalities

All participants have successfully completed the demographics questionnaire, the two tasks and the retrospective think-aloud session. It happened twice that a participant ran into trouble during a task. One participant thought at one moment that the wrong data was selected, which was caused by a mistake in the website. This made the participant go back to the calendar. Another participant did not use the dropdown menu to select a destination, which made that the destination was not correctly recognised. The participant had to go back and select the destination again.

As both mistakes did not happen on a moment in the experiment that had to do with dark patterns, the data of both participants was left in. If the parts where the participants were deviating from the task flow would have been cut out of the recorded time, this would not have made a difference in the outcome of the experiment. In the reporting of the results – unless mentioned differently – the unaltered results are presented.

The eye tracking software provided quality scores between 0 and 100. These indicate how much of the time the eyes of the participant were correctly recorded by the eye tracking device. The average quality scores of the eye tracking data lay between 90.1 and 99.3 for the various scenes in the control condition and between 90.3 and 98.9 for the experimental condition. Only the relevant scenes in which dark patterns were applied and that were hence relevant to the experiment are taken into account in this calculation. Those scores were considered as high enough to be acceptable for use in the analysis. A full overview can be found in appendix A.

4.2.2. Preparing the data

Quantitative data. For the quantitative analysis mainly data from the eye tracking software was used. For each participant various timestamps were exported from the iMotions software. Those included the duration of each 'scene' (i.e. a single page on the website in the task) and the total duration of each of the two tasks. Those data were connected to the data from the survey and the conditions the participant was in (experimental or control).

This data collection was extended with the data from the Areas of Interest. For each of the dark patterns in the experimental website an AOI was drawn around the place were the dark pattern was located. An overview of the locations of the AOIs is shown in appendix B. For each of those AOIs the iMotions software calculated the data for the available metrics for each of the participants. Those data were extracted from the data files of the iMotions project and brought together in a combined data sheet, labelled with the participant identification code, condition and dark pattern instance they belonged to.

Qualitative data. The qualitative data of this research consisted on one hand of the heat maps and gaze patterns that were formed by the way participants looked at the screen and on the other hand the data from the retrospective think-aloud sessions.

As an addition to the quantitative metrics of AOIs heatmaps were used to analyse the way the participants looked at the webpages. These were generated by the iMotions software for each single page on the website. After separate heatmaps were generated for the control condition and the experimental condition they were exported.

The gaze patterns were analysed separately by watching the gaze movements of each participant in the gaze recordings. In those videos dots and lines visualise fixations and saccades respectively, based on which recurring patterns could be counted.

The audio recordings of the retrospective think-aloud sessions were transcribed within a few days after each sessions. Quotes that are used in texts in the following sections sometimes had to be translated from Dutch to English.

4.3. Time based results

This subsection discusses the time based results of the experiment. First the overall duration of each of the tasks is discussed, after which the completion time for scenes with a subtask in which dark patterns were applied are reported. Table 5 contains a quick overview of all statistical calculations of the time based metrics. The green cells indicate that the reported times (in milliseconds) are normally distributed (based on a Shapiro-Wilk Test). If for both conditions the data is normally distributed the independent samples t-test is used. Before this an F-test of equality of variances is done to determine whether equal variances can be assumed. In the column with the test result the yellow cells indicate that equal variances were assumed in the t-test. For non-normally distributed data the Mann-Whitney U-test is used. The last column shows whether the results indicate a significant difference between the two conditions (based on $\alpha = 0.05$).

	Metric	Mean (SD) control	Mean (SD) experiment	Test results	Significant
Times (in milliseconds)	Overall duration task 1	63425.710 (21798.630)	79313.670 (25823.850)	U=30, p=0.234	
	Overall duration task 2	79113.000 (16733.480)	69572.140 (21232.850)	t(11)=-0.888, p=0.394	
	Duration of the checkout process task 1	13320 (8267.91)	23450.5 (7395.291)	t(11)=-2.310, p=0.041	**Significant result**
	Duration of closing cookie pop-up	5151.833 (1101.709)	6629.714 (6738.036)	U=26, p=0.534	
	Time to close the confirmshaming popup (green one)	3877.667 (484.2609)	4854.714 (2083.501)	t(6.748)=-1.203, p=0.269	

Table 5 Brief overview of the time-based metrics of the experiment.

4.3.1. Overall duration

The overall duration involves the full duration of the task from the moment the participants starts the task until they close the screen at the end of the task. In task 1 it took the participants in the experimental condition on average a bit longer (79313.67 ms, $\sigma = 25823.85$ ms) to complete the task than the participants in the control condition (63425.71 ms, $\sigma = 21798.63$ ms). This difference of almost 16 seconds turned out to be not significant ($U = 30$, $p = 0.234$).

For task 2 the experimental condition took the participants on average a bit less time to complete (69572.14 ms, $\sigma = 21232.85$ ms) than the control condition (79113 ms, $\sigma = 16733.48$ ms). The difference was however small (about 9,5 seconds) and not significant ($t(11) = -0.888$, $p = 0.394$).

A repairment can be carried out for the two participants that took for the earlier described reasons a wrong route on the website. This can be done by cutting out the time from the moment something goes wrong up till the moment that they are back on track. Doing this makes the average times go down a bit, but the differences stay small and not significant. As their deviating task flow did not occur on pages were dark patterns were used this did not influence other parts of the analysis.

4.3.2. Duration of checkout process task 1

In the checkout process of task 1 (train task) the participants were in the experimental condition faced with three instances of dark patterns (*false hierarchy*, *sneak into basket* and a *trick question*, shown in Figure 14). A comparison in the duration between the control and experimental condition was made based on the exported timestamps for this scene.

In the experimental condition this step took the participants about 23450.5 ms (σ = 7395.291) to complete, which was significantly longer than the 13320 ms (σ = 8267.91 ms) it took participants in the control condition ($t(11) = -2.310$, $p = 0.041$).

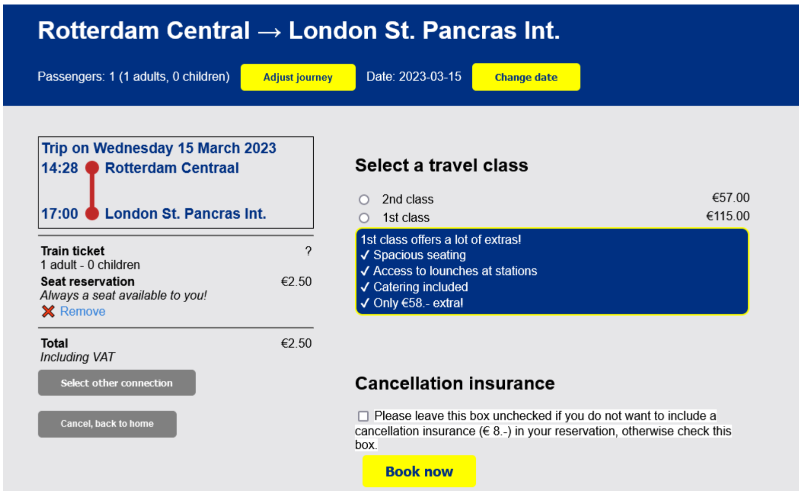

Figure 14 Screenshot of 'Checkout process task 1', with *false hierarchy*, *sneak into basket* and *trick question*.

4.3.3. Duration of closing cookie pop-up

At the start of task 2 (hotel task) the participant had to either accept or decline cookies in a cookie consent pop-up (shown in Figure 15). In this pop-up the dark pattern *preselection* was used in the experimental condition. The time it took the participants to enter their choice and close the pop-up differed between the two version by about a second: 5151.833 ms (σ = 1101.709) for the control condition, compared to 6629.714 (σ = 6738.036) for the experimental condition. No significant difference was found between the two versions ($U = 26$, $p = 0.534$).

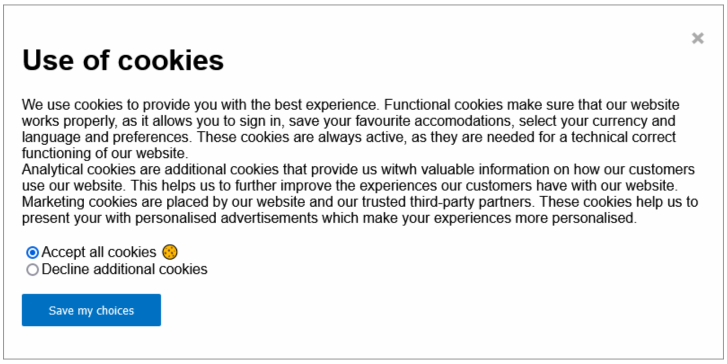

FIGURE 15 COOKIE POP-UP IN TASK 2 WITH *PRESELECTION*.

4.3.4. Duration of closing the nagging pop-up

Before the results were displayed in the experimental condition at the hotel booking website (task 2) a pop-up with an advertisement was shown, as an implementation of *nagging* (shown in Figure 16). It took on average 2638.429 ms (σ = 945.9617) before the participants in this condition closed the screen. The control condition did not include a *nagging* pop-up, so no comparison could be made for this part of the experiment.

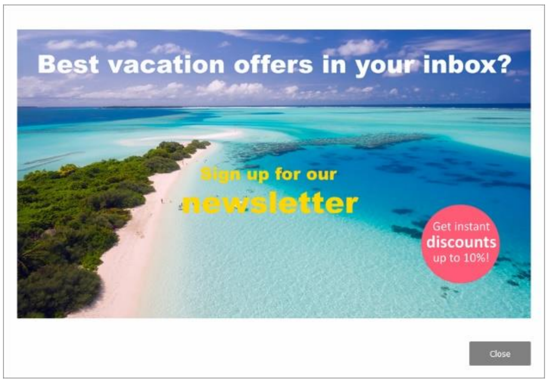

FIGURE 16 *NAGGING* POP-UP IN TASK 2.

4.3.5. Durations of closing the confirmshaming pop-up

In the final steps of task 2 (hotel task) the participants had to choose a room type, which was interrupted by a pop-up that tried to persuade the participant into booking a more luxurious type of room (shown in Figure 17). The difference between the experimental and control condition however was the text of the closing button, which was either "Close" in the control condition, or "No, thanks, I would rather skip this offer and have a simpler experience" in the experimental condition. The latter option is a form of *confirmshaming*. In the experimental condition it took on average 4854.714 ms (σ = 2083.501) before the pop-up was closed, which was a bit longer than in the control condition, in which it took on average 3877.667 ms (σ = 484.261). The difference however was not significant ($t(6.748) = -1.203$, $p = 0.269$).

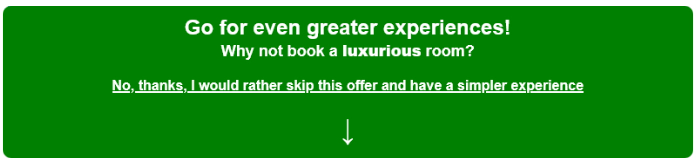

FIGURE 17 POP-UP WITH *CONFIRMSHAMING* IN IT.

4.4. Results dark pattern AOIs

The previous section discussed the duration of (parts of) the tasks, which already give some indication on how the use of dark patterns influences the time spent on a webpage. It is however possible to zoom in further on the application of dark patterns and analyse them in more details with the help of Areas of Interest (AOIs). This is especially useful for the page on which multiple dark patterns were applied. In this section, for each of the dark pattern AOIs, first the quantitative results will be discussed (measurements based on AOI data) and subsequently the qualitative data (heatmaps and remarks from the retrospective think-aloud session).

The way the AOIs were drawn are shown in appendix B. This is usually around the place where the dark pattern is applied. In the case of for example *sneak into basket* it is drawn around the shopping basket. The *low-stock/high-demand messages* however form exception: here the AOI was drawn over all hotel results shown on the page, which included the *low-stock and high-demand messages*, but also the other information that is shown. This was done because if AOIs had been drawn only around the *low-stock/high-demand messages* the comparison in the control version would only be against empty spaces (as the messages are absent there). It is hence more informative to draw the AOIs around a larger area.

For each of the AOIs discussed in the following subsections seven metrics were taken into account. First the *gaze dwell count* was analysed, which is the number of times the gaze of the participant entered the AOI. The *hit time AOI* metric indicated how long it took before the gaze of the participant entered the AOI for the first time from the moment the scene started. The *gaze dwell time* is a measurement in milliseconds indicating the total time the participant's gaze was in the AOI. Next to gazes also *fixations* are analysed, which are the moments when the gaze of the participant was located on the same position for a longer time. The *fix count* is the number of times the participant fixated within the AOI. The *TTFF (Time To first Fixation) AOI* indicates how long it took in milliseconds before the participant fixated for the first time in the AOI from the start of the scene. The *fixation dwell time* measures how long the participants fixated in the AOI. Finally the *number of mouse clicks* in each of the AOIs was analysed.

Table 6 gives a summarised overview of the three metrics that were found the most relevant in the analysis, which were gaze dwell time, fix count and fixation dwell time. Appendix C contains the complete version of the table, which contains all the metrics. The green cells indicate that the reported data are normally distributed (based on a Shapiro-Wilk Test). If for both conditions the data is normally distributed the independent samples t-test is used. Before this an F-test of equality of variances is done to determine whether equal variances can be assumed. In the column with the test result the yellow cells indicate that equal variances were assumed in the t-test. For non-normally distributed data the Mann-Whitney U-test is used. The last column shows whether the results indicate a significant difference between the two conditions (based on $\alpha = 0.05$). Figure 18 and Figure 19 give a visual overview of the dwell times and fix counts respectively, based on the data shown in the table.

In the following subsections the quantitative results will be discussed, together with the corresponding heatmap and comments form the retrospective think-aloud session.

Dark pattern	Metric	Mean (SD) control	Mean (SD) experiment	Test results	Significant
Sneak into basket	Gaze dwell time	1042.648 (588.837)	4945.413 (3197.011)	t(5.291)=-2.948, p=0.030	Significant result
	Fix count	4.143 (1.574)	15.833 (9.923)	t(5,213)=-2.854, p=0.034	Significant result
	Fixation dwell time	899.824 (357.140)	4483.446 (2918.452)	t(5,1285)=-2.989, p=0.030	Significant result
Trick question	Gaze dwell time	1285.456 (895.362)	8681.51 (4971.926)	t(5.278)=-3.594, p=0.014	Significant result
	Fix count	4.857 (3.132)	32.333 (18.533)	t(5.245)=-3.588, p=0.015	Significant result
	Fixation dwell time	1040.592 (577.794)	7438.561 (4631.942)	t(5.134)=-3.361, p=0.019	Significant result
False hierarchy	Gaze dwell time	2168.27 (1120.072)	4411.58 (1059.712)	t(11)=-3.689, p=0.004	Significant result
	Fix count	7.857 (4.947)	12.5 (3.564)	t(11)=-1.908, p=0.083	
	Fixation dwell time	1835.786 (751.7396)	3855.73 (1162.147)	t(11)=-3.781, p=0.003	Significant result
Preselection	Gaze dwell time	3215.449 (1092.521)	2054.993 (897.041)	t(11)=2.105, p=0.059	
	Fix count	9.833 (4.167)	8.143 (3.934)	U=27, p=0.424	
	Fixation dwell time	3099.809 (1042.5)	1828.546 (1008.153)	t(11)=2.231, p=0.047	Significant result
Low-stock/High demand	Gaze dwell time	10519.71 (2099.788)	9739.77 (4038.955)	t(11)=0.425, p=0.679	
	Fix count	37.333 (3.933)	32.571 (16.791)	t(6.759)=0.727, p=0.491	
	Fixation dwell time	8262.134 (2235.399)	6619.535 (3438.002)	t(11)=0.100, p=0.339	
Confirm-shaming	Gaze dwell time	3532.285 (746.5318)	3864.955 (1880.369)	t(11)=-0.405, p=0.693	
	Fix count	10.667 (2.066)	13.714 (5.438)	t(11)=-1.289, p=0.224	
	Fixation dwell time	3005.33 (1019.45)	2497.786 (1564.245)	U=30, p=0.234	

TABLE 6 BRIEF OVERVIEW OF THE MOST RELEVANT METRICS IN THE ANALYSIS. TIMES (GAZE AND FIXATION DWELL TIME) ARE REPORTED IN MILLISECONDS, FIX COUNT IN NUMBER OF FIXATIONS.

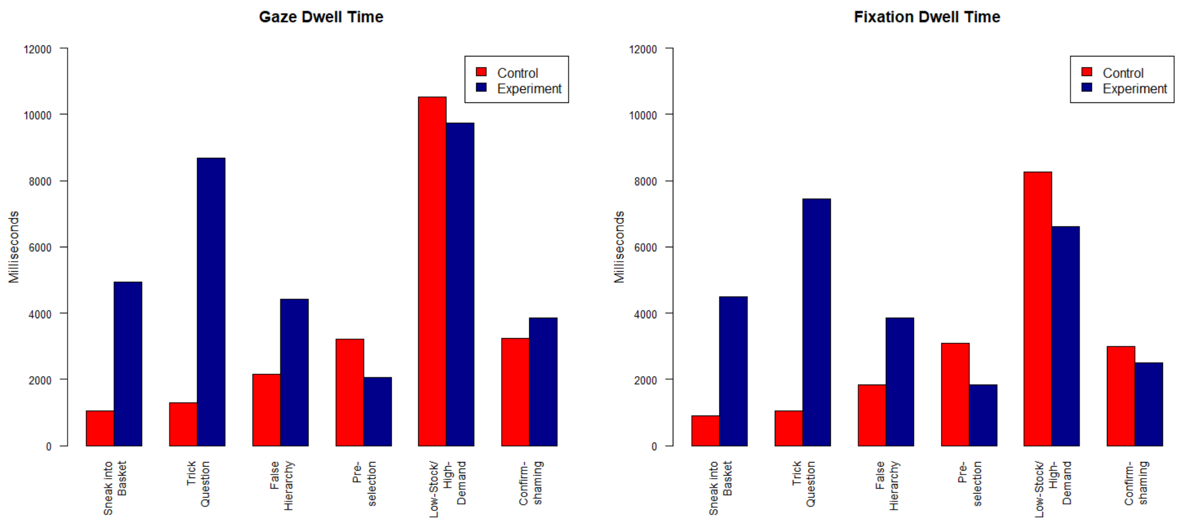

FIGURE 18 BAR CHARTS SHOWING THE AVERAGE GAZE AND FIXATION DWELL TIME FOR THE DARK PATTERNS ANALYSED.

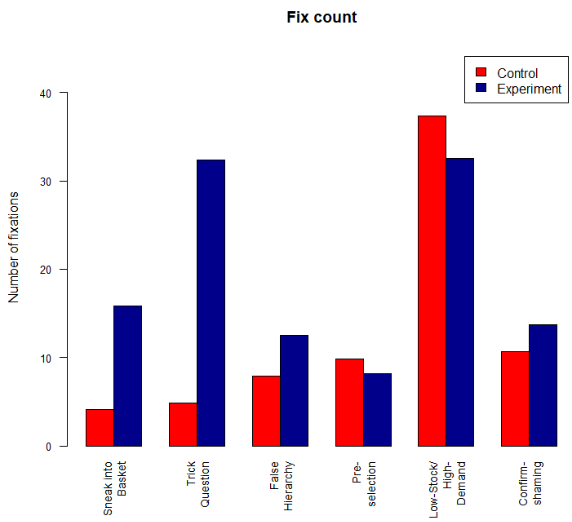

FIGURE 19 BAR CHART SHOWING THE AVERAGE FIX COUNT FOR THE DARK PATTERNS ANALYSED.

4.4.1. Sneak into basket

The pattern *sneak into basket* was included in task 1. In the experimental version an extra product was added to the purchase overview of the participant, in contrast to the control version where only the selected journey was part of the overview.

Of the 6 participants that completed task 1 in the experimental condition only 2 did actually remove the extra item from the purchase overview (33.3%). In the task description the participants were instructed to book a train journey only and that they did not need to have any extras included.

Gaze and fixation metrics. The gaze dwell time was in the experimental condition (4945.413 ms, $\sigma = 3197.011$ ms) significantly longer compared to the control condition (1042.648, $\sigma = 588.837$) ($t(5.291) = -2.948$, $p = 0.030$). This also applied to the fixation dwell time, which was in the experimental condition (4483.446 ms, $\sigma = 2918.452$ ms) likewise significantly longer than the control condition (899.824 ms, $\sigma = 357.140$)

(t(5,1285) = -2.989, p = 0.030). The fixation count was in the experimental condition (15.833, σ = 9.923) more than three times as high as in the control condition (4.143, σ = 1.574), which was also a significant difference (t(5,213) = -2.854, p = 0.034). For the other metrics no significant differences were found.

Heatmap. Figure 20 shows a comparison of the heatmaps for the purchase overview of task 1. It shows that in general there was a more intense gaze in the experimental condition – in which an item was sneaked in the purchase overview – compared to the control version in which this was not the case.

FIGURE 20 HEATMAPS OF THE CONTROL VERSION (LEFT) AND EXPERIMENTAL VERSION (RIGHT) FOR THE *SNEAK INTO BASKET* PATTERN.

Comments. Not many participants specifically commented on the *sneak into basket* method, someone noted indeed to have missed it after being pointed to the presence of this pattern.

4.4.2. Trick question

A *trick question* asking whether the participant needed a cancellation insurance was included in task 1. In the experimental condition the question was phrased in a convoluted way ("*Please leave this box unchecked if you do not want to include a cancellation insurance (€ 8.-) in your reservation, otherwise check this box.*") compared to the much simpler wording in the control condition ("*Include a cancellation insurance (€ 8.-) in your reservation*").

Gaze and fixation metrics. For the AOI around the *trick question* the gaze dwell time was significantly longer in the experimental condition (4411.580 ms, σ = 1059.712 ms) compared to the control condition (2168.27 ms, σ = 1120.072 ms) (t(11) = -3.689, p = 0.004). This was also the case for the fixation dwell time, where the number of milliseconds was again higher in the experimental condition (7438.561 ms, σ = 4631.942) than in the control condition (1040.592 ms, σ = 577.794). This was again a significant difference (t(5.134) = -3.361, p = 0.019). The fix count was also higher in the experimental condition (32.333, σ = 18.533) than in the control condition (4.857, σ = 3.132), which was also a significant difference (t(5.245) = -3.588, p = 0.015). The other metrics did not show a significant difference between the two conditions for this AOI.

Heatmap. The heatmaps in Figure 21 show the differences between the two versions. In the control version there is much less focus on the question and there is also a noteworthy lack of gazing to the end of the question. The experimental version gets much more visual attention that is also more spread out over the whole line.

FIGURE 21 HEATMAPS OF THE CONTROL VERSION (LEFT) AND EXPERIMENTAL VERSION (RIGHT) FOR THE *TRICK QUESTION* PATTERN.

Comments. A lot of participants in the experimental condition commented on the *trick question*. They for example mentioned that it was hard to understand what it was saying, such as a participant saying: "I had to read it five times to get what it was saying", or "I did not really understand what it was saying". Other participants also said that it was confusing, which sometimes made that it took more time to complete the task: "I thought the wording was so strange that I lingered here for a while to check whether I was doing it right", or "…and then I had to carefully read this, whether I had to do something or not, in the end I though it was not the case".

4.4.3. False hierarchy

The choice between first and second class for the train journey was in the experimental condition part of a *false hierarchy*: the first class option was promoted over the second class option. In the control option this was left out.

Gaze and fixation metrics. The mean gaze dwell time was 2168.270 ms (σ = 1120.072 ms) in the control condition and 4411.58 ms (σ = 1059.712 ms) in the experimental condition, which means that the gaze dwell time was significantly longer in the experimental condition ($t(11) = -3.689$, $p = 0.004$). Also the fixation dwell time was longer in the experimental condition (3855.73 ms, σ = 1162.147 ms) compared to the control condition (1835.786 ms, σ = 751.7396 ms), which was also a significant difference ($t(11) = -3.781$, $p = 0.003$). For the other metrics no significant differences were found between the two versions.

Heatmap. Figure 22 shows the heatmaps of the two version. Both show a strong focus of the participants on the radio buttons that they had to use to make their choice. Also a slight focus is visible on the component trying to persuade the user into choosing a first class ticket.

FIGURE 22 HEATMAPS OF THE CONTROL VERSION (LEFT) AND EXPERIMENTAL VERSION (RIGHT) FOR THE *FALSE HIERARCHY* PATTERN.

Comments. Some participants commented on the *false hierarchy* between the options for first and second class: "[it said] that I didn't have first class… all the things I would mis without first class". Another participant said that it did not have any effect: "I didn't read that, because I just wanted second class".

4.4.4. Preselection

The *preselection* pattern was used in the cookie pop-up that was shown at the start of task 2. In the experimental condition the option “Accept all cookies” was preselected, whereas in the control condition participants had to select this option or “Decline additional cookies” themselves.

The acceptance rates were not high in the sample: in the control version 1 of the 6 participants accepted the cookies and in the experimental version 2 of the 7 participants did so.

Gaze and fixation metrics. The AOI to analyse *preselection* was drawn over the area in which the choice has to be made (the radio button and confirmation button) as this is the place where the pattern itself is actually applied. The data shows that the fixation dwell time in the experimental version (1828.546 ms, σ = 1008.153 ms) is significantly shorter than in the control version (3099.809 ms, σ = 1042.5 ms) (t(11) = 2.231, p = 0.047), but the other metrics do not show significant differences here as well.

Heatmap. Figure 23 shows the heatmaps of the cookie pop-ups with and without the *preselection* applied. A strong focus is visible around the radio buttons where the participants had to select one of the options. Except for the slightly more spread out focus on the text above the buttons – which is not part of the actual dark pattern – in the experimental version no other noteworthy differences are visible.

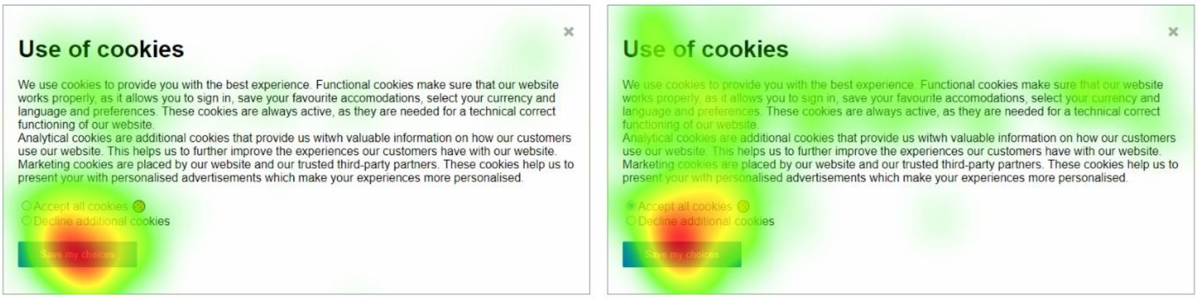

FIGURE 23 HEATMAPS OF THE CONTROL VERSION (LEFT) AND EXPERIMENTAL VERSION (RIGHT) FOR THE *PRESELECTION* PATTERN.

Comments. Some participants commented on the fact that they had to make a choice in the pop-up (“Cookies, I honestly thought this was already quite annoying”), or on the choice they made in the end (“Yeah, I though I will just accept it”). No-one commented on the fact that a choice was already preselected.

4.4.5. Nagging

A pop-up based on the *nagging* pattern was shown before the search results in task 2 appeared. As the appearance of this pop-up is the dark pattern itself, having a control condition was not possible for this pattern, except for leaving it out at all.

Gaze and fixation metrics. The average gaze dwell time for the *nagging* pop-up was 2350.236 ms (σ = 873.932 ms) and the average fixation dwell time 1613.973 ms (σ =

430.293). For the latter one it has to be noted that this metric is based on one participant less, as this person did not fixate at all on the pop-up content.

Heatmap. Figure 24 shows the heatmap of the *nagging* pop-up. The "close"-button on the bottom right of the pop-up gets the most visual attention.

FIGURE 24 HEATMAP OF THE *NAGGING* PATTERN.

Comments. Some participants commented on the sudden appearance of the pop-up ("I hadn't expected this one") or that it was annoying ("It got in the way"). Most of the participants however only mentioned that they closed it very quickly ("I thought: I will just close it, I don't need it", "I just closed it really quickly, I thought: I don't need all of that", "No, I am not interested, just quickly reading what it was saying").

4.4.6. Low-stock/High-demand messages

The *low-stock and high-demand messages* were placed next to some of the accommodations in the list on the hotel booking website in task 2. In the control condition these were left out.

Gaze and fixation metrics. The AOI for this dark pattern was placed over the centre of all the boxes with results, to see whether participants paid more visual attention to the information of the hotels when *low-stock and high-demand messages* were included. The AOIs were not drawn only around the messages themselves, as this would have been less informative in combination with the control condition. A comparison would then only have been possible by drawing AOIs over empty parts of the page, at which none of the participants would have paid any attention. None of the metrics showed a significant difference between the two versions.

Heatmap. The heatmaps for the moments where this pattern occurs are shown in Figure 25. It shows the participants scanning the accommodations and a strong visual focus on the button they had to click in the end. Messages belonging to the pattern are included in the second, third and fourth box. According to the heatmap the messages are sometimes seen, but no strong focus is visible.

FIGURE 25 HEATMAPS OF THE CONTROL VERSION (LEFT) AND EXPERIMENTAL VERSION (RIGHT) FOR THE *LOW-STOCK/HIGH-DEMAND MESSAGES* PATTERN.

Comments. None of the participants commented on this pattern. Two participants however mentioned that they had seen these messages after it was mentioned as an example of a dark pattern during the explanation of the goal of the experiment at the end.

4.4.7. Confirmshaming

The pattern *confirmshaming* was applied in a pop-up that appeared when the participants had to select a type of room (a standard room according to the instructions), which tried to persuade them to book a more expensive room.

Gaze and fixation metrics. An AOI was drawn around the pop-up that appeared once the participant moved their cursor over the area in which a room type could be selected. The AOI was only active while the pop-up was visible. For none of the metrics of the AOI around the pop-up a significant difference between the two conditions was found.

Heatmap. Figure 26 contains the heatmaps for the two conditions. On the heatmap for the experimental version a somewhat stronger and more spread out focus is visible.

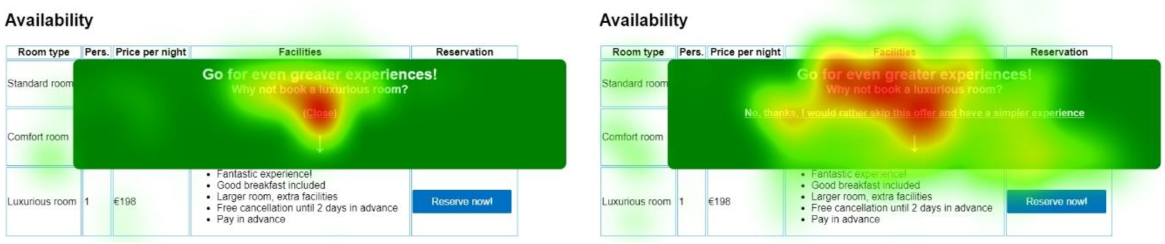

FIGURE 26 HEATMAPS OF THE CONTROL VERSION (LEFT) AND EXPERIMENTAL VERSION (RIGHT) FOR THE *CONFIRMSHAMING* PATTERN.

Comments. A lot of the comments were about the appearance of the box itself. Someone commented for example: "I found it annoying". People also commented on the fact that it was hard to close it: "I was like how do I close this thing?", or "I was wondering why I had to read the whole sentence before I knew I could close it there". Some people did not read the text on the pop-up ("I didn't actually… like gonna read it", "Then I just thought: no, I don't want that (…) I usually close these things very fast"), while someone else said: "So then I had to read this a few times".

4.5. Visual patterns

This section describes interesting or noteworthy visual gaze patterns that were (in a qualitative way) discovered during the analysis. First the method of analysing that was used is introduced and subsequently the results are reported.

The previous sections have presented the results using statistical calculations, visualisation and user comments. This section will add to this by looking at the visual gaze patterns. There are many visualisation and analysis techniques available that provide insight in eye tracking data (Raschke, Blascheck, & Burch, 2014; Blascheck, Kurzhals, Raschke, Weiskopf, & Ertl, 2017). For this research there is a specific interest at gazing behaviour on a small scale (for example within a certain area), instead of analysing on a larger scale (for example between multiple areas or for a whole page), which is where most the aforementioned techniques focus on.

Hence another way of analysing the visual gaze patterns had to be found. In this case the eye tracking recordings of all participants for each scene – a small part of a task – were watched. In those videos the places where the participant fixated were shown as dots, the movements between those fixations (saccades) were shown as lines. An illustrative example of this visualisation is shown in Figure 27.

FIGURE 27 EXAMPLE OF VISUALISATION OF GAZES OF A PARTICIPANT. DOTS REPRESENT A FIXATION, LINES REPRESENT SACCADES.

While watching those videos noteworthy patterns were noted. No categories were set on beforehand, those were created while watching the videos. If a new category came up late in the process the earlier videos would be rewatched. This way of categorising is inspired on the coding technique based on the Grounded Theory that is often used for analysing interviews (Deterding & Waters, 2011).

In the following subsections for each of the seven dark patterns that were part of the experiment the visual gaze patterns that were found and how often they were found are reported. Table 7 gives an overview of the gaze patterns found. In this table for each dark pattern the gaze patterns that were found are mentioned, for which in the last column the total number of participants at which this gaze pattern was seen is reported. This number is in relation to the total number of participants for the specific task in the experimental condition. The first three dark patterns are from the first task, in which 6 participants were in the experimental condition, the other dark patterns are from the second task, in which 7 participants were in the experimental condition.

Dark pattern	**Gaze pattern**	**Number of participants at which the gaze pattern is seen** *(in relation to the total number of participants in the experimental condition)*
Sneak into basket	Fixating on the Sneak into Bakset	5/6
Trick question	Fixating before reading	2/6
	Regressions	4/6
	Starting over	4/6
	Stops reading halfway through	1/6
	No clear reading pattern	1/6
False hierarchy	Fixating on false hierarchy	6/6
	Reading the bullet points	2/6
Preselection	Fixating on the option labels	6/7
	No fixation on the option labels.	1/7
Nagging	Reading content	2/7
	Scanning content	4/7
Low-stock/high-demand messages	Fixating on low-stock/high-demand messages	7/7
Confirmshaming	*No patterns found*	

TABLE 7 OVERVIEW OF GAZE PATTERNS FOUND.

4.5.1. Sneak into basket

For the analysis of the *sneak into basket* pattern the recordings of the checkout screen in task 1 (Figure 9) were watched. For this pattern only one category of visual patterns was identified: *Fixating on the Sneak into Basket*. This label was assigned to a recording if there was at least one moment on which the participant was fixating on the extra item (a seat reservation) that was added to their reservation. This was the case for 5 of the 6 participants.

4.5.2. Trick question

For the analysis of the *trick question* also the recordings of the checkout screen of task 1 (Figure 9) were watched. A variety of different kinds of behaviour was seen around this pattern. One pattern was *Fixating before reading*, which means that a participant fixated at least once on the *trick question*, but did not start reading and moved to another part of the page. During the reading three gaze behavioural patterns were seen: *regressions*, which involves participants that did not read from the beginning to the end of the sentence, but were moving repeatedly moving to a point earlier in the sentence with their gaze (Matlin & Farmer, 2017, pp. 90-91). An example of this is shown in Figure 28. *Starting over* implies that someone was reading the *trick question* and at a certain point (either halfway or at the end) goes back to the start of the sentence to start reading again. There is also the option that someone *stops reading halfway through the sentence*, meaning that someone stops reading at a certain point in the sentence (not being (almost) at the end of the sentence) and starts looking at another element on the page. Finally it is also possible that there is *no clear reading pattern* visible at all, which is assigned to participants for who their gaze pattern does not clearly show any reading behaviour at all.

FIGURE 28 ON THE LEFT FIXATIONS 5-7 SHOW READING, IN THE MIDDLE FIXATIONS 8 AND 9 SHOW A REGRESSION, AS THEY APPEAR EARLIER IN THE SENTENCE THAN 7. ON THE RIGHT: A LONGER REGULAR READING PATTERN (WITHOUT REGRESSION).

Fixating before reading occurred to 2 of the 6 participants. While reading there is 1 participant that *stops reading halfway through the sentence*. For the majority of the participants – 4 out of 6 – *regressions* are visible while reading. All those participants also *started over* at a certain point at least once. 1 participant *stops reading halfway through the sentence*. For another participant *no clear reading pattern* is visible at all. For those last two mentioned participants no other labels were assigned than the ones mentioned.

To compare this to the control condition: for 6 of the 7 participants who saw the regular text next to the checkbox *no clear reading pattern* was visible, only some loose fixations or saccades going over the sentence. Only for 1 participant *regression* was visible, this participant also *starts over* and lies with respect to the gaze patterns relatively close to the participants in the experimental condition.

4.5.3. False hierarchy

To analyse the *false hierarchy* the recordings of the checkout screen of task 1 (Figure 9) were used as well. Two gaze behavioural patterns were identified: *fixating on false hierarchy*, meaning that the participant fixated on the box causing the false hierarchy; and *reading the bullet points*, meaning that reading behaviour was visible on at least one of the bullet points in the box belonging to the *false hierarchy* pattern.

All 6 participants showed to some extent *fixations on false hierarchy*. 2 of the participants were marked as *reading the bullet points*.

4.5.4. Preselection

For the analysis of the *preselection* pattern the recordings of the cookie pop-up of task 2 (Figure 10) were used. Each participant was labelled with one of the following categories: *fixating on the option labels* and *no fixation on the option labels*. The first option indicates that at least one fixation is seen within the region around the radio buttons and the latter indicating that this is not case.

6 of the 7 participants were *fixating on the option labels*, one was not doing this and hence categorised as *no fixation on the option labels.*

4.5.5. Nagging

To analyse *nagging* the pop-up shown halfway through task 2 (Figure 11) was used. By doing this two behavioural patterns were identified: *reading content* and *scanning content*. The first one indicating that reading behaviour was seen in at least part of the advertisement in the pop-up, the latter one indicating that only some loose fixations on various places within the content of the pop-up were seen. Figure 29 shows examples of both those categories.

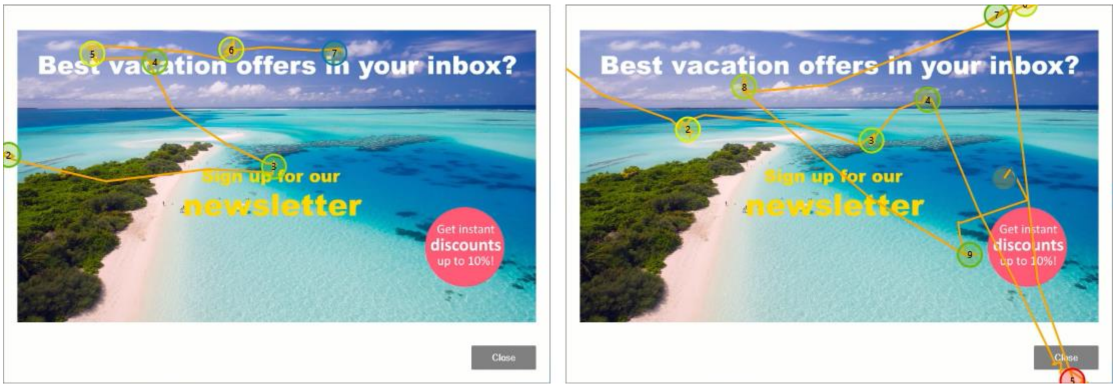

FIGURE 29 EXAMPLE OF *READING CONTENT* ON THE RIGHT AND *SCANNING CONTENT* ON THE LEFT.

Of the 7 participants, 2 showed at least some signs of *reading content*. The other 4 all did fixate at the pop-up, but were only *scanning content*. The last participant that is not in any of the two categories makes no fixations, but quickly moves over the line of text at the top with only saccades. As a sidenote it is important to keep in mind that the average time it took before the pop-up was closed was around 2.6 seconds (as discussed in section 4.3.4.). This means that even for the participants that were reading the content, this did in general only involve quickly reading a sentence and not reading a full paragraph of text.

4.5.6. Low-stock/High-demand messages

For the *low-stock/high-demand messages* the recordings of the screen in which those were added to the search results for the hotels (Figure 12) was used. The only pattern that could be identified here was *fixating on low-stock/high-demand messages*. This behaviour was seen for all participants.

4.5.7. Confirmshaming

For *confirmshaming* no relevant gaze patterns were discovered.

5. Discussion

This section discusses the results presented in the previous section and places them in a broader perspective.

5.1. Data, validity and generalisability

The goal of this research was to investigate gaze patterns and time duration in the context of dark patterns. The data for this research have been collected in an eye tracking experiment with 13 participants. All participants have successfully completed the full experiment, including the pre-experiment questionnaire and the post-experiment retrospective think-aloud session. Two participants deviated from the task flow, which made that they spent more time on the task. This did however not have an effect on the outcome of the analysis of the total time and as the deviations did not happen on a moment were dark patterns were presented it had no effect on the analysis of the dark patterns.

Apart from those two cases no abnormalities did occur during the experiment. The quality of the data overall was good, as for each of the scenes in which dark patterns occurred (or their counterpart in the control condition) the averaged eye tracking data quality score was between 92 en 98 (out of 100).

The generalisability of the experiment is relatively quite low. This has to do with the sample size and representativity of it. A sample size of 13 is enough to show some basic effects, to find significant results, and to do qualitative analysis, but is not enough to draw strong conclusions about the general populations. This has also to do with the limited diversity of the sample. In the sample namely almost 70% of the population was female, which is a large over-representation. Also more than 60% of the sample was aged between 21 and 30 years old, underrepresenting other age groups. Finally, almost 85% of the sample had an educational background on university master level, underrepresenting other educational backgrounds. This all together means that the conclusions from this research are indicative only and cannot be extended to the whole population without further research.

While interpreting the results it is also important to keep in mind that all the results are based on fixations and saccades of the eye. This gives valuable insight in what someone is doing, but when someone is fixating on a point it does not guarantee that this person is also paying attention let alone understands what is shown at that point.

5.2. Time

The first research question in this research was "*How do dark patterns affect the time users need to complete a task?*". To answer this question the time-related metrics of the experiment can be used. The analysis showed that the overall durations of the tasks did not differ between the condition with and without dark patterns. The subparts of the task were analysed on a page level. This showed durations that were not significantly different between the two conditions, expect for the checkout screen of task 1. On this screen participants spent more time when dark patterns were included compared to the version without dark patterns. This indicates that the dark patterns present on this screen (*sneak into basket*, *false hierarchy* and *trick question*) make that people need more time to complete such a process.

The fact that this part of the task took longer in the experimental version is partly reflected in the total time of the task, as this was about 15 seconds longer than the control version, but not a significant effect. It might be possible that this is partly caused by the small sample size.

As the screen that showed a significant difference in time contained multiple patterns (*sneak into basket*, *false hierarchy* and *trick question*) it is not possible to say which dark pattern or combination of dark patterns caused this effect. The data of the Areas of Interest (AOIs) however offers some more insight in the specific effects that individual pattern have, which is discussed in the next subsection.

5.3. AOIs

The analysis related to the Areas of Interest (AOIs) that were drawn around instances of dark patterns in the experiment provides more insight in the time spent looking at those. During the analysis around all instances of dark patterns an AOI was drawn, after which the metrics of this AOI were compared between the control and experimental condition.

For three dark patterns – *sneak into basket*, *false hierarchy* and *trick question* – it was found that the participants spent significantly more time looking at it compared to the same area in the control condition. This conclusion can be drawn based on the fact that for all these three patterns the gaze dwell time and fixation dwell time were significantly higher in the experimental version, indicating that the gazes and fixations of the participants were longer in total when looking at the dark patterns. For *sneak into basket* and *trick question* the fixation count was also significantly higher, indicating that participants not only did look longer but also had more fixations in the experimental version.

Based on this it can be concluded that – within the limitations of the experiment – the dark patterns *sneak into basket*, *false hierarchy* and *trick question* cost people more time to deal with compared to a similar interface without those dark patterns.

For the other dark patterns in the experiment (*preselection*, *nagging*, *low-stock*/*high-demand messages* and *confirmshaming*) no such results were found. For the *preselection* pattern – used in the cookie pop-up – however an opposite results was found: considering the AOI drawn around the radio buttons and confirmation button, the participants in the experimental version had a significantly lower fixation dwell time compared to the control version in which the dark pattern were implemented. In this case the results suggest that when the pattern *preselection* is applied people spend less visual attention on the area where their choice can be entered. A logical explanation for this would be that in the experimental version 2 of the 7 participants (about 28%) did not change the preselected option. This means that they also did not need to look very long in the selection area, whereas all 6 participants in the control condition had to select one of the two options themselves, which is likely to cause a higher average fixation dwell time. The total time it took before participants closed the pop-up did not significantly differ between the control and experimental condition. This is probably caused by the variation that is present between each participant (some might for example read some more text in the pop-up than others, some might look at the website or the task description where others do not) which mitigates the effects a dark pattern can have on time spent on a page. A larger sample might provide more insight into this.

5.4. Gaze patterns

The second research question in this research was "*What visual gaze patterns can be seen around applications of dark patterns?*". To answer this question the gaze patterns that were seen around the instances of dark patterns have been analysed.

For the pattern *sneak into basket* it turned out that all participants except for one fixated at least once on the extra product that was added to their reservation. However only 2 of the 6 participants removed the product from the reservation. It cannot with certainty be explained why there are 3 participants who have seen the extra product but did not remove it. A possible explanation is that the participants were not actively processing the information they saw in enough detail to realise that there was something odd going on. Another possibility is that the participants did not know that they were supposed to remove this extra product, as they were not explicitly instructed to do so, although it was stated that no extra products had to be included in the booking. This can then be seen as a consequence of the fact that the task was done in an experimental setting: the participants might have watched the price better or reacted differently if it had been a real purchase.

Around the *trick question* a variety of gaze patterns was seen. 4 of the 6 participants showed a form of jumping with their eyes back and forth while reading (moving along the sentence but then suddenly moving back). Those so-called regressions occur in general more in poor readers compared to good readers (Matlin & Farmer, 2017, pp. 90-91). The cause of those regressions is often that people realise that they have not fully understood the sentence that they were reading (Matlin & Farmer, 2017, pp. 90-91; Rayner, 1998) and hence more regressions are made if the text someone is reading is complex (Booth & Weger, 2013).

The gaze patterns seen around the *trick questions* in the experiment – in which regression in reading was seen often around the *trick question* – can hence be explained with this theory: the complexity of the *trick question* makes that the participants have a remarkably high number of regressions.

The dark pattern *preselection* was applied to a cookie pop-up in the second task. Here only one participant did not fixate on the option labels, whereas the rest of the participants did so. It is however hard to make statements about how the application of dark patterns has influenced this. The duration and options the participants selected are more informative for this, which were discussed in the previous subsections.

For the *false hierarchy* it turned out that all participants fixated at least once on the *false hierarchy* choice, but only 2 of the 6 participants showed clear reading behaviour. This means that only a part of the participants were actually perceiving (parts of the) context of the nudging texts, while the majority only had a quick look and further ignored the text.

The *nagging* pop-up that was shown halfway through task 2 was labelled with two types of gaze patterns. 2 of the 7 participants were (quickly) reading the content of the pop-up, while the 4 others were only scanning (i.e. fixating on some loose points) the content and a final participant only moves over the text with saccades. This shows that there is difference between participants who were paying more and less attention to what was in the pop-up, but as discussed in the section considering the time related data all participants closed the pop-up in a very short amount of time. This shows even more than the gaze patterns that the participants had a quick reflex of immediately closing the pop-up.

For the *low-stock and high-demand messages* the gaze pattern analysis showed that all participants fixated at least once on one of the *low-stock or high-demand messages*.

6. Conclusion

This thesis research has tried to provide more insight into how people look at dark patterns and how this influences the time they spend on a page. The research questions that this research tried to answer were: h*ow do dark patterns affect the time users need to complete a task?* (RQ1) and *what visual gaze patterns can be seen around applications of dark patterns?* (RQ2).

To answer these question an experiment has been setup with two tasks in which seven instances of dark patterns were included. All participants (N = 13) completed the two tasks (of which one was in the control condition) while being eye tracked. Before the start of the experiment the participants completed a short demographics questionnaire and they were asked some questions afterwards while rewatching the eye tracking video (retrospective think-aloud).

This section will summarise the findings and based on what follows from this answer the research questions. It will subsequently discuss the implications the research has and in the end discuss its limitations and the possibilities for future research in this field.

6.1. Research conclusions

The first research question focusses on how dark patterns might influence the time users need to complete a task. Based on the logged times of the eye tracking software and the data of the Areas of Interest (AOI) and within the boundaries and limitations of the experiment, it can be concluded that the dark pattern *false hierarchy*, *trick question* and *sneak into basket* are likely to cost users more time than a comparable situation in which these dark patterns are not applied.

This conclusion is based on the fact that the data from the AOIs around those dark patterns indicate that the participants spent more time looking at them, compared to the control condition. The participants also spent more time on the screen in which those dark patterns were found.

For the pattern *preselection* it was found that the participants spent *less* time looking at the radio buttons in a cookie pop-up in the experiment, compared to the same pop-up in which no *preselection* was applied. This is likely caused by the fact that when *preselection* is applied, an action from the user is not necessary (in contrast to a choice where each user has to select one of the options) and hence some users might hardly look at the options, if they are satisfied with the selected option. In the experiment the total time spent before closing the cookie pop-up did not differ.

For the other three dark patterns in the experiment no significant differences in the time spent on the page nor the time spent looking at the specific dark pattern instances were found.

The second research question focused on visual gaze patterns. This provided some preliminary insight in how people look at and behave around dark patterns on websites. For *sneak into basket* it was seen that almost each participant fixated on the extra product that was 'sneaked' into to the reservation of the participant, but only a small number of people actually removed the extra product. This can mean that people did not notice this extra product, but can also be caused by the experimental setting (i.e. not realising what is going on and what the experimental instructions asked from them). For the *trick question* a lot of regressions were seen: people moving their eyes back to earlier parts of the sentence. According to what is known about eye behaviour while reading this happens more often in difficult texts, which would be a reasonable explanation in the case of a *trick question*. For the *false hierarchy* the gaze pattern analysis showed that all participants fixated at least once on the box causing the false hierarchy, but only two of them actually read (parts of) what was in there. For *nagging* part of the participants quickly read some parts of the content, whereas the other part only made some quick fixations without clear reading patterns.

6.2. Implications

As mentioned earlier, a lot of research is currently taking place on the topic of dark patterns. However as far as known not much is yet known about how much time dark patterns users cost and how to look at instances of dark patterns. This research contributes to the knowledge on those two topics.

The newly introduced categorisation that distinguishes timebound and local dark patterns offers a new way of looking at different types of dark patterns, in addition to the existing taxonomies and categorisations.

The results of the experiment provide more insight in how users interact with interfaces in which dark patterns are used. This adds to the total knowledge about dark patterns, and can for example be taken into account in legal questions, as it gives preliminary indications on which types of dark patterns do disrupt the user the most. This connects to various recent researches that are trying to come up with a way to classify the severity of dark patterns, such as Cara (2019) and Van Nimwegen, Bergman & Akdag (2022).

On a more practical side especially the findings with regard to the time needed for an action are interesting. It can provide owners and designers of apps with new insights in how dark patterns influence the user flow and user experience of their users. It can also help them in weighing again the pros and cons of dark pattern usages. Dark patterns are usually implemented with a certain goal in mind (such as making people buy more products), but a possible side effect might be that people spend more time or attention to these elements, which might not always be intended.

6.3. Limitations

There are various limitations that can be mentioned considering this research. The most important drawback is probably the fact that the results from the research cannot easily be generalised. This is because the research consisted of a lab experiment in which participants might not have acted exactly in the same way as they would have done if they were not participating in an experiment. It was sometimes also indicated by participants in the retrospective think-aloud session that they were for example paying more attention to certain things or double-checking what they were doing, because the knew they were part of an experiment. The low generalisability is also caused by the fact that the sample is not very representative: females, young people and people with a university background are overrepresented. As certain groups (for example older people in contrast to young people) might react differently to dark patterns, no solid conclusions can be drawn for the whole population.

A larger group of participant would not only have made the research more representative, but might have also provided more results. Currently for some of the seven dark patterns significant differences have been found. With a larger sample more results might have shown up.

The experimental material generally functioned well and after the experiment was completed it was possible to analyse the data based on the data recorded. However, if the three dark patterns that were implemented on one page (*sneak into basket*, *trick question*, and *false hierarchy* at the end of task 1) had been placed on different pages it would have been easier to make statements about the total time spent on a page and the influence dark patterns have on it. With this setup it was only possible to state this based on the combination of the three patterns, supplemented with the data of the AOIs, that still made it possible to analyse each pattern separately.

6.4. Future work

In future work it would be important to first of all take the drawbacks of the current research into account and resolve them, for example by making sure there is a larger and more diverse sample. Next to that future research can also exist of conducting a comparable experiment, but with a larger number of dark patterns involved. Now only seven dark patterns were researched, which can be extended by researching more patterns. If there is for more patterns knowledge on the effects they have on the time they cost users and the gaze patterns they evoke, more general conclusions on how dark pattern influence users can be drawn.

Especially the findings and results on the gaze patterns that were found are preliminary. It is clear that dark patterns can evoke certain gaze patterns in the users of digital systems, but further research is needed in order to draw stronger conclusions. With a larger corpus of eye tracking data similarities, differences and patterns for certain types of dark patterns might be found.

Another related line of research that might benefit from further study using eye tracking is the (lack of) recognition of dark patterns by users (dark pattern blindness). Research to this topic has for example been done by Bhoot, Shinde and Mishra (2020) and Di Geronimo et al. (2020). As both studies show that dark patterns are not always noticed by users, eye tracking research can help in showing whether people miss those dark patterns because they have not seen them at all or that they did see them but did not realise that they were being tricked.

A final suggestion for future research would be to investigate the opposite of dark patterns – bright patterns – with a comparing study. Bright patterns are suggested as an ethical way of creating user interfaces, as they prioritise the values and goals of the user instead of the goals of the system owner (Graßl, Schraffenberger, Zuiderveen Borgesius, & Buijzen, 2021; Sandhaus, 2023). A comparing eye tracking study between dark and bright patterns can provide further insight in how persuasive interfaces work and whether their intention (user or owner prioritisation) makes a difference in how users interact with them.

References

Acquisti, A., Adjerid, I., Balebako, R., Cranor, L. F., Komanduri, S., Leon, P. G., . . . Wilson, S. (2018, May). Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online. *ACM Computing Surveys, 50*(3), 1-41. doi:10.1145/3054926

Bhatnagar, A., Sinha, A. P., & Sen, A. (2019). Role of navigational ability in website visit duration. *European Journal of Marketing, 53*(5), 972-988. doi:10.1108/EJM-10-2017-0719

Bhoot, A. M., Shinde, M. A., & Mishra, W. P. (2020). Towards the Identification of Dark Patterns: An Analysis Based on End-User Reactions. *Proceedings of the 11th Indian Conference on Human-Computer Interaction*, (pp. 24-33). doi:10.1145/3429290.3429293

Blascheck, T., Kurzhals, K., Raschke, M., Weiskopf, D., & Ertl, T. (2017, February 9). Visualization of Eye Tracking Data: A Taxonomy and Survey. *Computer Graphics Forum, 36*(8), 260-284. doi:10.1111/cgf.13079

Bongard-Blanchy, K., Rossi, A., Rivas, S., Doublet, S., Koenig, V., & Lenzini, G. (2021). "I am Definitely Manipulated, Even When I am Aware of it. It's Ridiculous!" - Dark Patterns from the End-User Perspective. *Designing Interactive Systems Conference 2021 (DIS '21)* (pp. 1-14). Virtual Event: Association for Computing Machinery. doi:10.1145/3461778.3462086

Booth, R., & Weger, U. (2013). The function of regressions in reading: Backward eye movements allow rereading. *Memory & Cognition, 41*, 82-97. doi:10.3758/s13421-012-0244-y

Bösch, C., Erb, B., Kargl, F., Kopp, H., & Pfattheicher, S. (2016). Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns. *Proc. Priv. Enhancing Technol., 2016*(4), 237-254. doi:10.1515/popets-2016-0038

Brignull, H. (n.d.). *Deceptive Design - Types of deceptive design*. Retrieved from Deceptive Design: https://www.deceptive.design/types

Brignull, H. (n.d.). *Deceptive Design - user interfaces crafted to trick you*. Retrieved from Deceptive Design: https://www.deceptive.design/

Cara, C. (2019, February). Dark patterns in the media: A systematic review. *Network Intelligence Studies, 7*(14), 105-113.

Cho, H., Powell, D., Pichon, A., Kuhns, L. M., Garofalo, R., & Schnall, R. (2019). Eye-tracking retrospective think-aloud as a novel approach for a usability evaluation. *International Journal of Medical Informatics*(129), 366-373. doi:10.1016/j.ijmedinf.2019.07.010

Choi, J., Seo, J., & Lee, G. (2009). Analysis of Web Usage Patterns in Consideration of Various Contextual Factors. *Proceedings of The 7th International Workshop on Intelligent Techniques for Web Personalization & Recommender Systems*, (pp. 63-74). Pasadena, California, USA.

Conti, G., & Sobiesk, E. (2010). Malicious interface design: exploiting the user. *WWW '10: Proceedings of the 19th international conference on World wide web* (pp. 271-280). Raleigh (North Carolina): Association for Computing Machinery. doi:10.1145/1772690.1772719

Deterding, N., & Waters, M. (2011). Flexible coding of in-depth interviews: A twenty-first-century approach. *Sociological methods & research, 50*(2), 708-739. doi:10.1177/0049124118799377

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. *CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-14). Honolulu: Association for Computing Machinery. doi:10.1145/3313831.3376600

Djamasbi, S., Siegel, M., Skorinko, J., & Tullis, T. (2011). Online viewing and aesthetic preferences of generation y and the baby boom generation: Testing user web site experience through eye tracking. *International Journal of Electronic Commerce, 15*(4), 121-158. doi:10.2753/JEC1086-4415150404

DSA. (2022, October 19). *Digital Services Act.* Retrieved from EUR-lex: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32022R2065&from=EN

Dumais, S. T., Buscher, G., & Cutrell, E. (2010). Individual differences in gaze patterns for web search. *Proceedings of the Third Symposium on Information Interaction in Context* (pp. 185-194). New York, NY, USA: Association for Computing Machinery. doi:10.1145/1840784.1840812

Eidelson, J. (2019, July 15). Instacart Hounds Workers to Take Jobs That Aren't Worth It. *Bloomberg*.

Fogg, B. J. (1999). Persuasive technologies. *Communications of the ACM, 42*(5), 26-29.

GDPR. (2016, April 27). *GDPR*. Retrieved from EUR-lex: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN#d1e1888-1-1

Goldberg, J., & Kotval, X. (1999). Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics, 24*(6), 631-645. doi:10.1016/S0169-8141(98)00068-7

Goldberg, J., Stimson, M., Lewenstein, M., Scott, N., & Wichansky, A. (2002). Eye tracking in web search tasks. *Proceedings of the 2002 symposium on Eye tracking research & applications* (pp. 51-58). New Orleans, Louisiana: Association for Computing Machinery. doi:10.1145/507079.507082

Graßl, P., Schraffenberger, H., Zuiderveen Borgesius, F., & Buijzen, M. (2021). Dark and Bright Patterns in Cookie Consent Requests. *Journal of Digital Social Research, 3*(1), 1-38. doi:10.33621/jdsr.v3i1.54

Gray, C. M., Kou, Y., Battles, B., Hoggatt, J., & Toombs, A. L. (2018). The Dark (Patterns) Side of UX Design. *CHI '18: Proceedings of the 2018 CHI Conference on Human*

Factors in Computing Systems (pp. 1-14). Montreal: Association for Computing Machinery. doi:10.1145/3173574.3174108

Gray, C., Santos, C., & Bielova, N. (2023). Towards a Preliminary Ontology of Dark Patterns Knowledge. *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-9). Hamburg, Germany: Association for Computing Machinery. doi:10.1145/3544549.3585676

Gunawan, J., Santos, C., & Kamara, I. (2022). Redress for Dark Patterns Privacy Harms? A Case Study on Consent Interaction. *Proceedings of the 2022 Symposium on Computer Science and Law* (pp. 181-194). Washington DC, USA: Association for Computing Machinery. doi:10.1145/3511265.3550448

Hernandez, A., & Resnick, M. L. (2013). Placement of call to action buttons for higher website conversion and acquisition: An eye tracking study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 1042-1046). Los Angeles, CA, USA: SAGE Publications. doi:10.1177/1541931213571232

Hofgesang, P. I. (2006). Relevance of Time Spent on Web Pages. *12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* Philadelphia, PA.

iMotions. (n.d.). *Eye Tracking.* Retrieved from iMotions.com: https://imotions.com/guides/eye-tracking/

Johnson, E. J., Bellman, S., & Lohse, G. L. (2022). Defaults, Framing and Privacy: Why Opting In-Opting Out. *Marketing Letters, 13*(1), 5-15. doi:10.1023/A:1015044207315

Kahneman, D. (2011). *Thinking, Fast and Slow.* New York: Farrar, Straus and Giroux.

Kelly, M. J. (2021, May 05). *What are deceptive design patterns?* Retrieved from Mozilla: https://blog.mozilla.org/en/internet-culture/mozilla-explains/deceptive-design-patterns/

Lam, H. (2008). A framework of interaction costs in information visualization. *IEEE transactions on visualization and computer graphics, 4*(6), 1149-1156. doi:10.1109/TVCG.2008.109

Luguri, J., & Strahilevitz, L. J. (2021). Shining a Light on Dark Patterns. *Journal of Legal Analysis, 13*(1), 43-109. doi:10.1093/jla/laaa006

Maier, M., & Harr, R. (2020). Dark design patterns: An end-user perspective. *Human Technology, 16*(2), 170-199. doi:10.17011/ht/urn.202008245641

Mathur, A., Acar, G., Friedman, M. J., Lucherini, E., Mayer, J., Chetty, M., & Narayanan, A. (2019, November). Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. *Proceedings of the ACM on Human-Computer Interaction, 3*(CSCW), 1-32. doi:10.1145/3359183

Mathur, A., Kshirsagar, M., & Mayer, J. (2021). What Makes a Dark Pattern... Dark?: Design Attributes, Normative Considerations, and Measurement Methods. *CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing*

Systems (pp. 1-18). Yokohama, Japan: Association for Computing Machinery. doi:10.1145/3411764.3445610

Matlin, M. W., & Farmer, T. A. (2017). *Cognition.* Hoboken (New Jersey): Wiley Custom.

McLaughlin, J. (2014, May 28). *9 iconic growth hacks tech companies used to boost their user bases*. Retrieved September 25, 2022, from The Next Web: https://thenextweb.com/news/9-iconic-growth-hacks-tech-companies-used-pump-user-base

Mittone, L., & Savadori, L. (2009). The Scarcity Bias. *Applied Psychology, 58*(3), 453-468. doi:10.1111/j.1464-0597.2009.00401.x

Nagy, I. K., & Gaspar-Papanek, C. (2009). User Behaviour Analysis Based on Time Spent on Web Pages. In I.-H. Ting, & H.-J. Wu (Eds.), *Web Mining Applications in E-Commerce and E-Services* (pp. 117-136). Heidelberg: Springer Berlin. doi:10.1007/978-3-540-88081-3_7

Narayanan, A., Mathur, A., Chetty, M., & Kshirsagar, M. (2020). Dark Patterns: Past, Present, and Future: The evolution of tricky user interfaces. *Queue, 18*(2), 67-92. doi:10.1145/3400899.3400901

Nielsen, J. (2006, April 16). *F-Shaped Pattern For Reading Web Content (original study)*. Retrieved from Nielsen Norman Group: https://www.nngroup.com/articles/f-shaped-pattern-reading-web-content-discovered/

Norman, D. (2013). *The Design of Everyday Things: Revised and Expanded Edition.* New York: 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. *CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-13). Honolulu: Association for Computing Machinery. doi:10.1145/3313831.3376321

Pernice, K. (2019, August 25). *Text Scanning Patterns: Eyetracking Evidence*. Retrieved from Nielsen Norman Group: https://www.nngroup.com/articles/text-scanning-patterns-eyetracking/

Raschke, M., Blascheck, T., & Burch, M. (2014). Visual Analysis of Eye Tracking Data. In W. Huang, *Handbook of Human Centric Visualization* (pp. 391-409). New York: Springer New York. doi:10.1007/978-1-4614-7485-2

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin, 124*(3), 372-422. doi:10.1037/0033-2909.124.3.372

Saffer, D. (2014). *Microinteractions - Designing with Details.* Sebastopol, CA, USA: O'Reilly Media, Inc.

Sandhaus, H. (2023). Promoting Bright Patterns [Preprint Workshop Paper]. *CHI '23 Workshop: Designing Technology and Policy Simultaneously, April 23, 2023, Hamburg, Germany*, (pp. 1-9). doi:10.48550/arXiv.2304.01157

Sharp, H., Rogers, Y., & Preece, J. (2019). *Interaction Design: beyond human-computer interaction* (5e ed.). Indianapolis: Wiley.

Sin, R., Harris, T., Nilsson, S., & Beck, T. (2022). Dark patterns in online shopping: do they work and can nudges help mitigate impulse buying? *Behavioural Public Policy*, 1-27. doi:10.1017/bpp.2022.11

Sinders, C. (2022, June 18). *What's In a Name? - Unpacking Dark Patterns versus Deceptive Design*. Retrieved from Medium.com: https://medium.com/@carolinesinders/whats-in-a-name-unpacking-dark-patterns-versus-deceptive-design-e96068627ec4

Soe, T., Nordberg, O., Guribye, F., & Slavkovik, M. (2020). Circumvention by design - dark patterns in cookie consent for online news outlets. *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society* (pp. 1-12). Tallinn, Estonia: Association for Computing Machinery. doi:10.1145/3419249.3420132

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge - Improving Decisions About Health, Wealth, and Happiness.* New Haven & London: Yale University Press.

Tversky, A., & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. In G. Wright, *Behavioral Decision Making* (pp. 453-458). Boston, MA: Springer. doi:10.1007/978-1-4613-2391-4_2

Van Nimwegen, C., Bergman, K., & Akdag, A. (2022). Shedding light on assessing Dark Patterns: Introducing the System Darkness Scale (SDS). *35th International BCS Human-Computer Interaction Conference (HCI2022)*, (pp. 1-10). doi:10.14236/ewic/HCI2022.7

Waldman, A. E. (2020, February). Cognitive biases, dark patterns, and the 'privacy paradox'. *Current Opinion in Psychology, 31*, 105-109. doi:10.1016/j.copsyc.2019.08.025

Williams, K. D., Cheung, C. K., & Choi, W. (2000). Cyberostracism: Effects of Being Ignored Over the Internet. *Journal of Personality and Social Psychology, 79*(5), 748-762. doi:10.1037//0022-3514.79.5.748

Zagal, J. P., Björk, S., & Lewis, C. (2013). Dark Patterns in the Design of Games. *Proceedings of the 8th International Conference on the Foundations of Digital Games* (pp. 39-46). Chania, Crete, Greece: Society for the Advancement of the Science of Digital Games.

Zuiderveen Borgesius, F., Kruikemeier, S., Boerman, S., & Helberger, N. (2017). Tracking Walls, Take-It-Or-Leave-It Choices, the GDPR, and the ePrivacy Regulation. *European Data Protection Law Review (EDPL), 3*(3), 353-368. doi:10.21552/edpl/2017/3/9

Appdenix

A. Eye tracking data quality

This appendix contains an overview of the average quality scores per scene for the control and experimental version, as well as the overall average. These averages are included in Table 8. Table 9 shows all the detailed scores per participant.

	Control	Experimental	Average
Checkout task 1	90.14	94.67	92.23
Cookie pop-up	99.33	90.29	94.46
Nagging	N.A.	98.86	98.86
Low-Stock/High-Demand messages	97.00	95.14	96
Confirmshaming	94.50	92.14	93.23

TABLE 8 AVERAGE QUALITY SCORES PER SCENE.

	Checkout task 1	Cookie pop-up task 2	Nagging task 2	Low-Stock/ High-Demand messages task 2	Confirm-shaming task 2	**Average overall**
Participants (Task 1 experimental, task 2 control)	100	100		95	85	**95.00**
	98	99		99	99	**98.75**
	99	100		97	87	**95.75**
	100	97		93	97	**96.75**
	98	100		99	99	**99.00**
	73	100		99	100	**93.00**
Participants (Task 2 experimental, task 1 control)	75	94	97	98	88	**90.40**
	89	39	100	82	96	**81.20**
	80	100	100	98	86	**92.80**
	100	100	100	100	100	**100.00**
	99	100	100	100	100	**99.80**
	100	99	95	88	86	**93.60**
	88	100	100	100	89	**95.40**
Average control	**90.14**	**99.33**	**N.A.**	**97.00**	**94.50**	
Average experimental	**94.67**	**90.29**	**98.86**	**95.14**	**92.14**	
Average overall	**92.23**	**94.46**	**98.86**	**96.00**	**93.23**	

TABLE 9 OVERVIEW OF QUALITY SCORES PER SCENE PER PARTICIPANT. THE BLUE COLOUR INDICATES THE CONTROL GROUP.

B. AOI locations

This appendix contains images of the scenes that show where the Areas of Interest (AOIs) are drawn. Figure 30 and Figure 31 show task 1, Figure 32, Figure 33, Figure 34, Figure 35, Figure 36, Figure 37 and Figure 38 show the scenes of task 2.

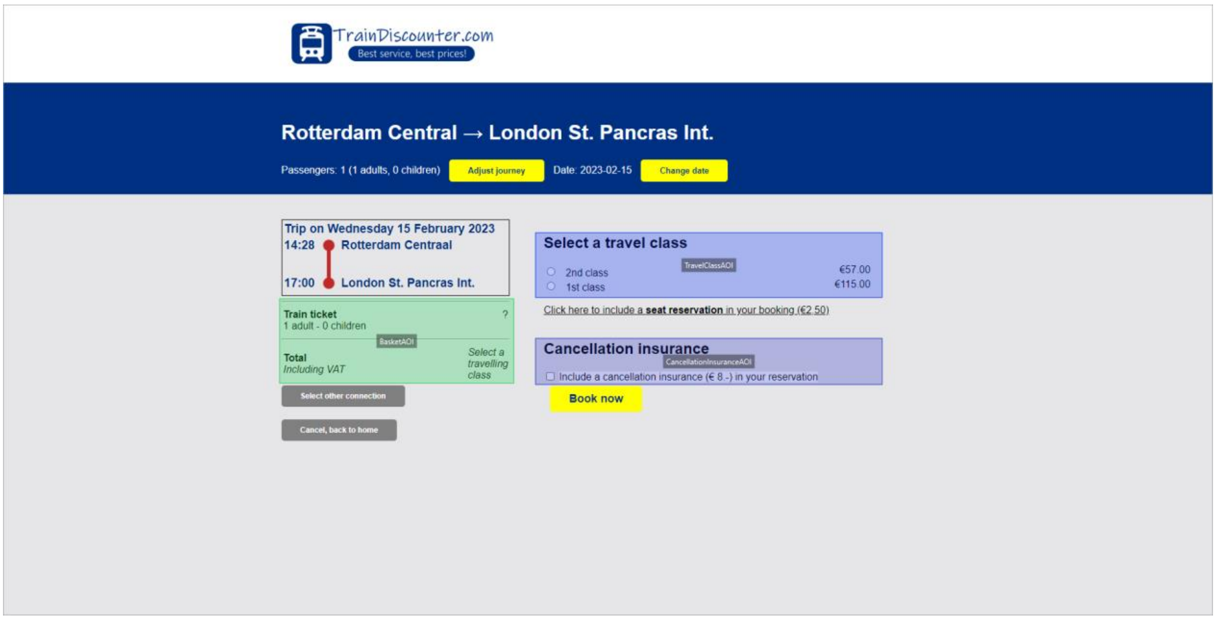

FIGURE 30 AOI DRAWINGS CONTROL CONDITION TASK 1 (*SNEAK INTO BASKET*, *FALSE HIERARCHY* AND *TRICK QUESTION*).

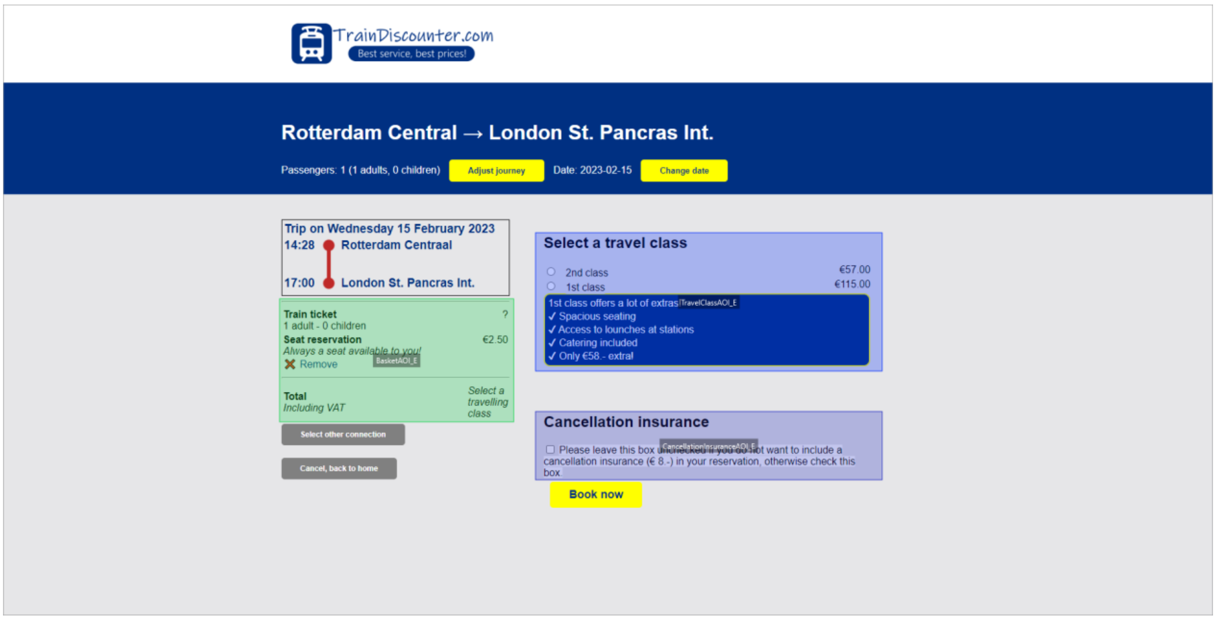

FIGURE 31 AOI DRAWINGS EXPERIMENTAL CONDITION TASK 1 (*SNEAK INTO BASKET*, *FALSE HIERARCHY* AND *TRICK QUESTION*).

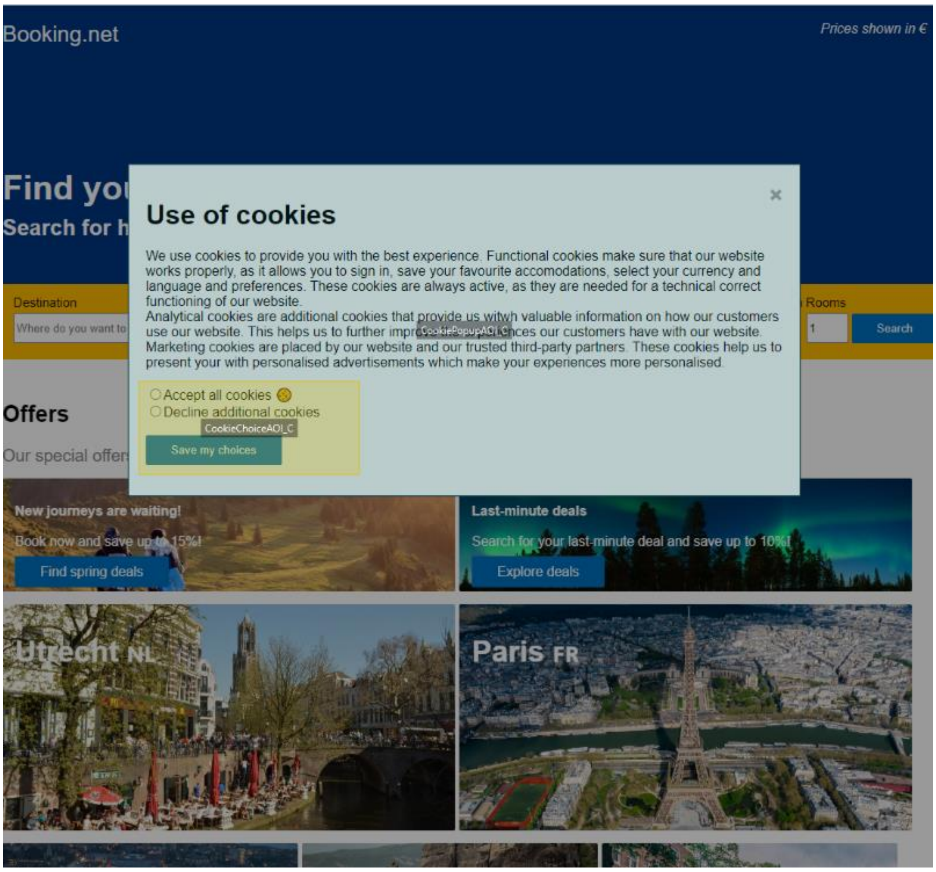

FIGURE 32 AOI DRAWING COOKIE POP-UP CONTROL CONDITION TASK 2 (*PRESELECTION*).

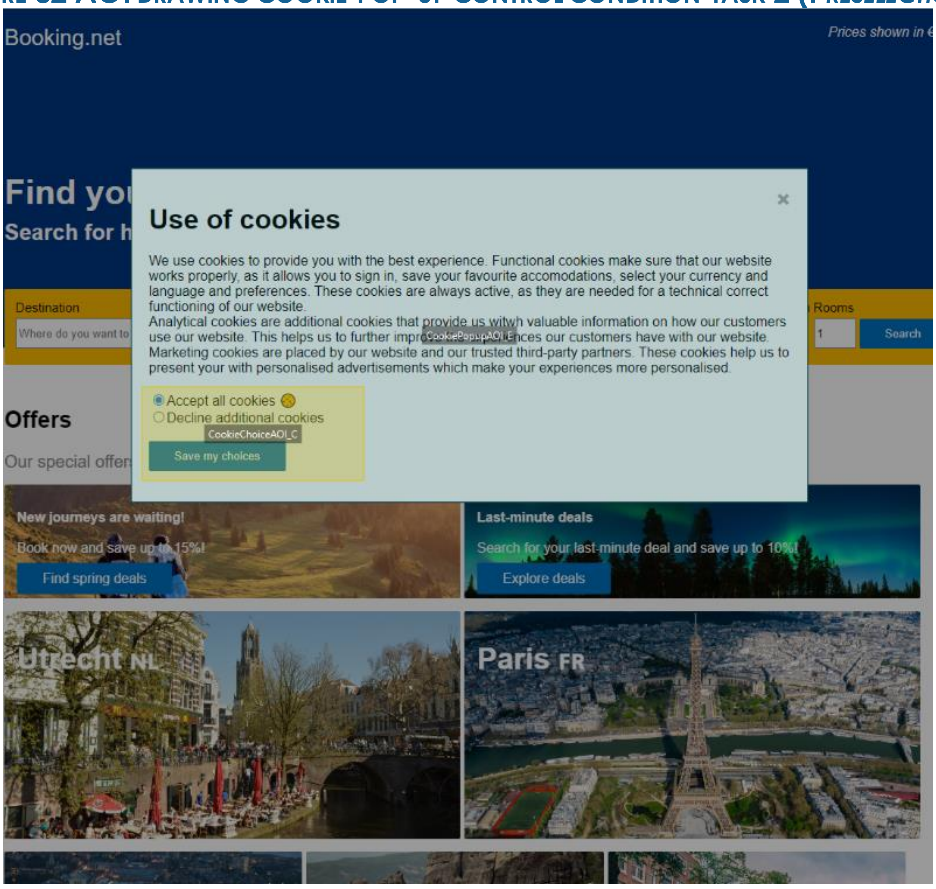

FIGURE 33 AOI DRAWING COOKIE POP-UP EXPERIMENTAL CONDITION TASK 2 (*PRESELECTION*).

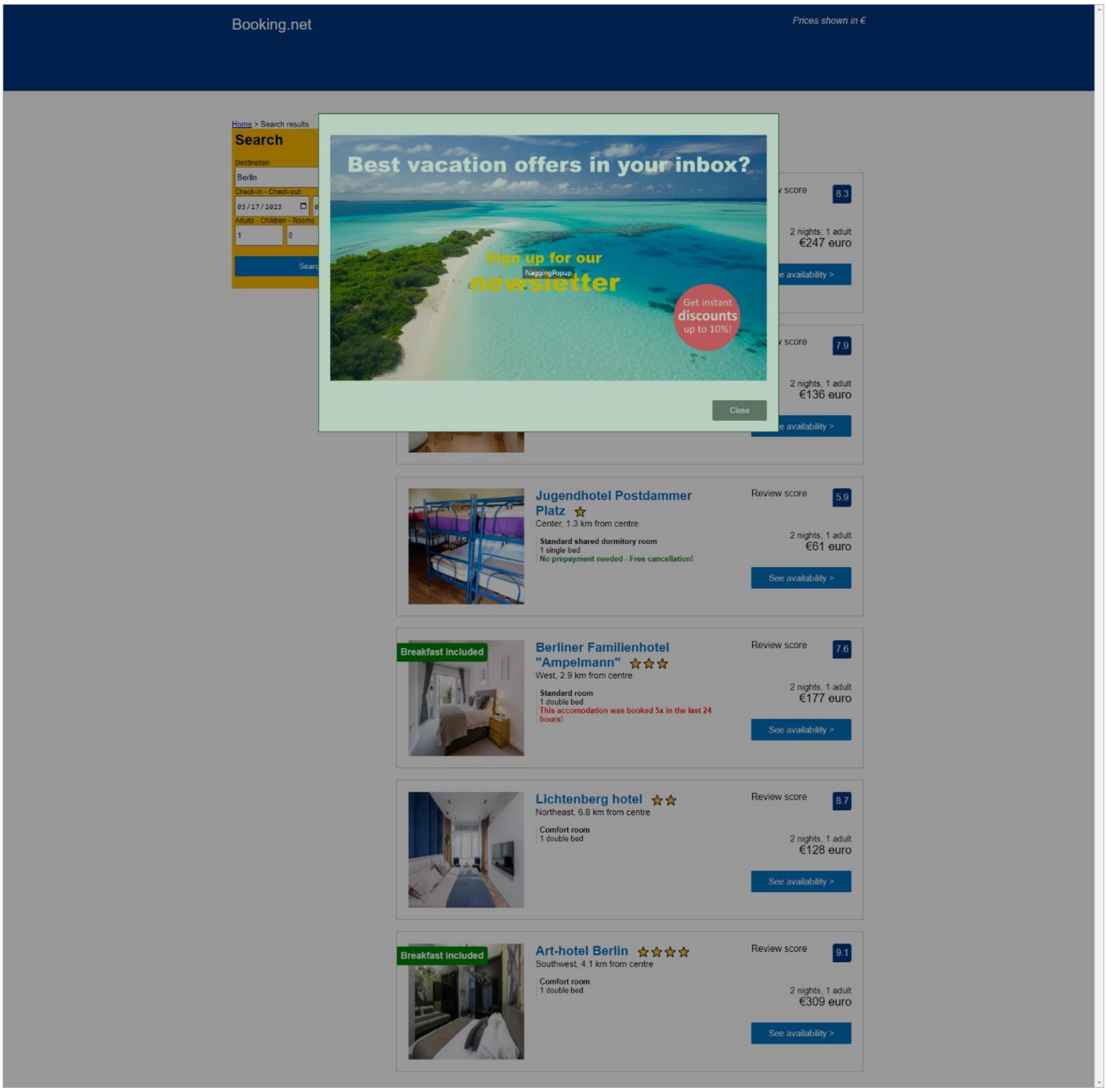

FIGURE 34 AOI DRAWINGS *NAGGING* POP-UP TASK 2.

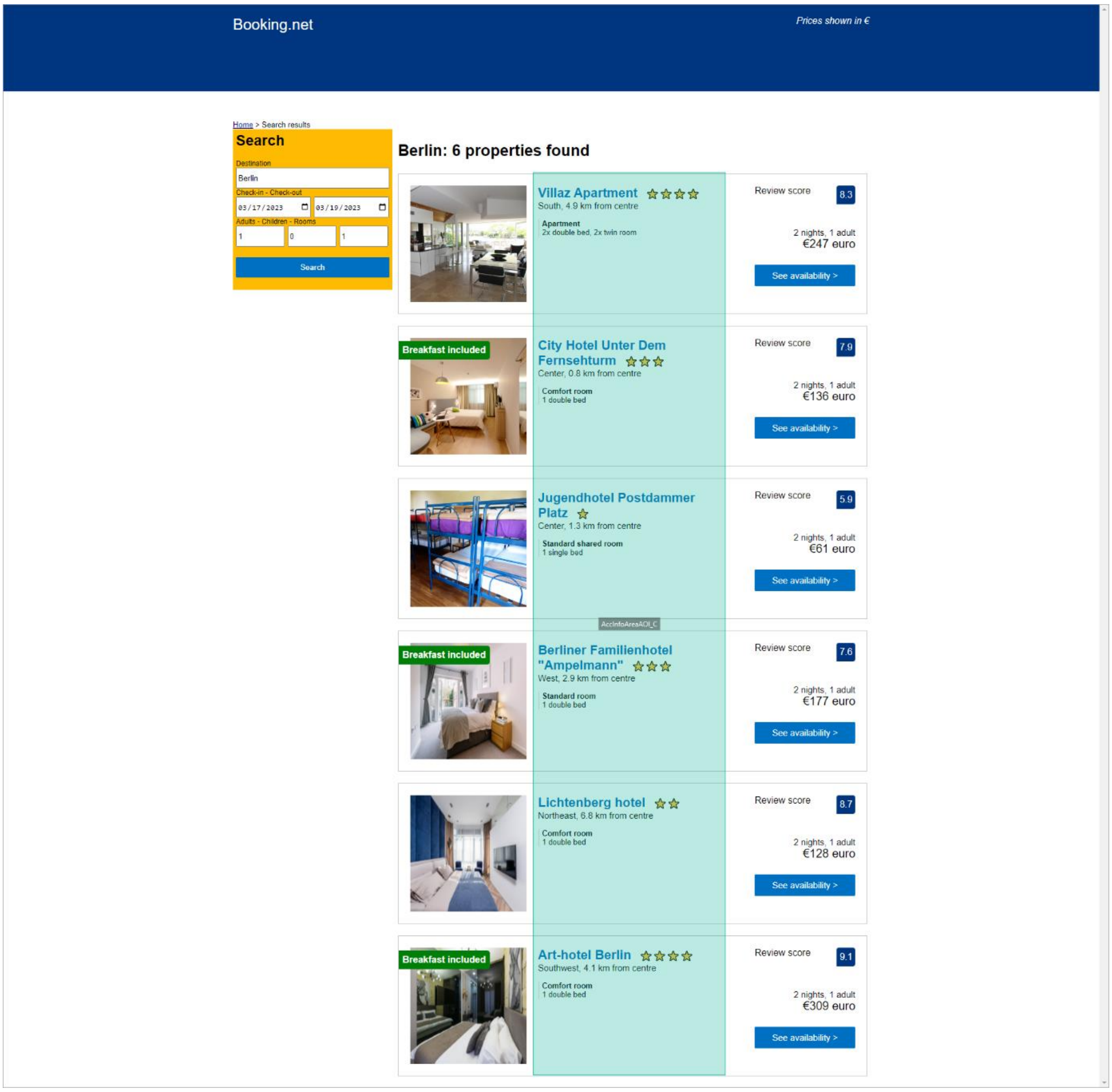

FIGURE 35 AOI DRAWING *LOW-STOCK/HIGH-DEMAND MESSAGES* CONTROL CONDITION TASK 2.

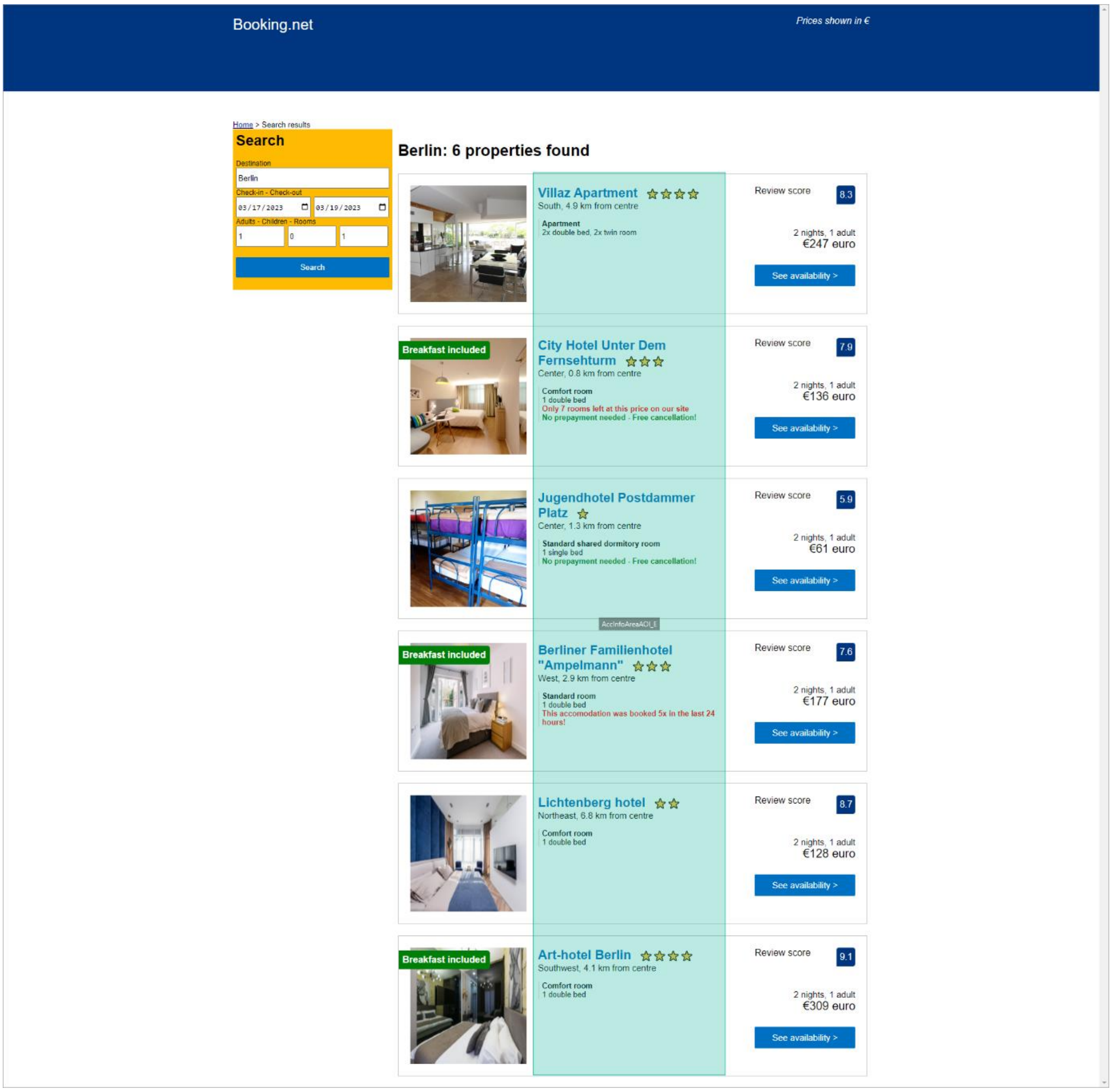

FIGURE 36 AOI DRAWING *LOW-STOCK/HIGH-DEMAND MESSAGES* EXPERIMENTAL CONDITION TASK 2.

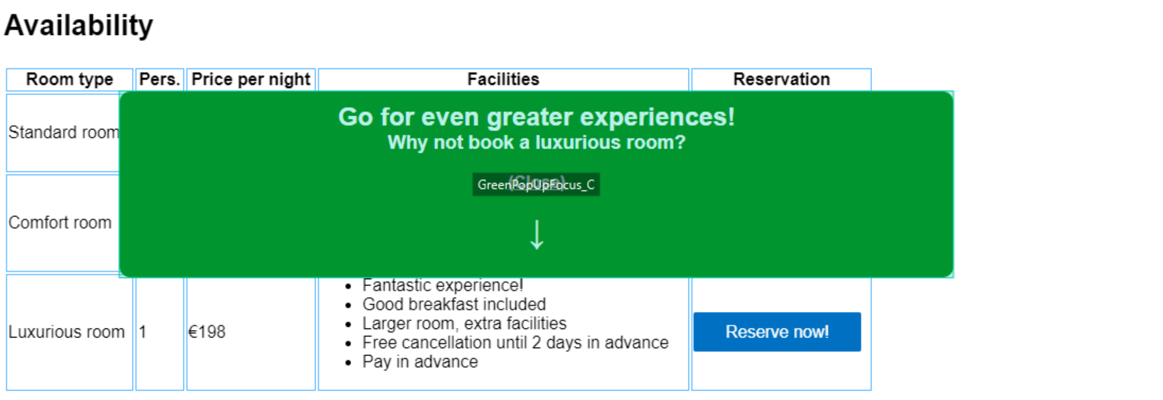

FIGURE 37 AOI DRAWING *CONFIRMSHAMING* CONTROL CONDITION TASK 2.

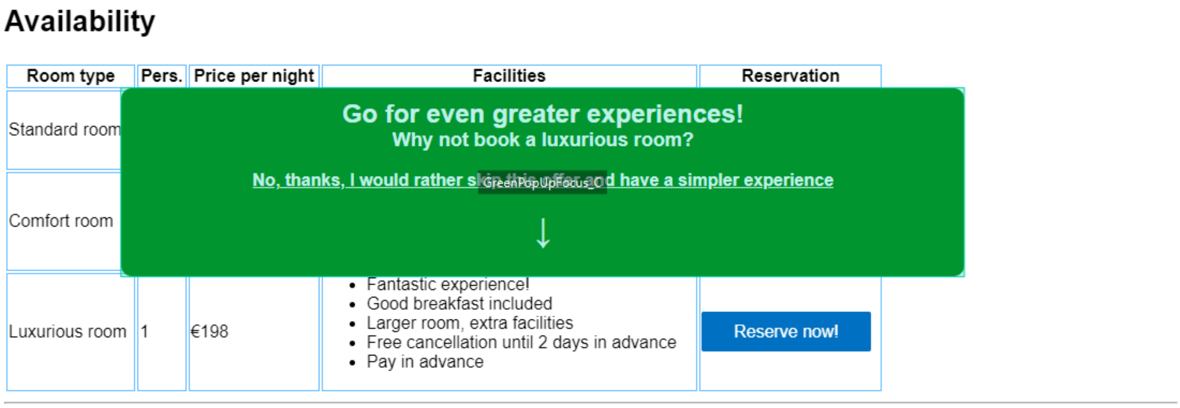

FIGURE 38 AOI DRAWING *CONFIRMSHAMING* EXPERIMENTAL CONDITION TASK 2.

C. Statistical overview

This appendix includes an overview of all statistical calculations that were made for the analysis of the experiment. The first column indicates for which dark pattern the AOI data was analysed. The second column shows which metric was used. The third and fourth column show the mean value and standard deviation for the control and experimental condition. The goal of the analysis is to see whether there is a statistically significant difference between those two. First of all the Shapiro Wilk test is done to see whether the data is normally distributed. If this is the case the column is coloured green. For normally distributed data an F-test of equality of variances is done to see whether equal variances can be assumed. With this information a t-test is done. A yellow background colour indicated that equal variances are assumed in the t-test. The last column shows whether the result found was significant. If the data is not normally distributed for a certain metric the Mann-Whitney U-tests was used (which does not require the F-test, hence this column is empty). The values in the table are rounded to three decimals.

(green) = Normally distributed (Shapiro Wilk test)

(yellow) = Equal variances (otherwise unequal variances/Welch test)

Dark pattern	Metric	Mean (SD) controle	Mean (SD) experiment	F-test	Test result	Significant
Sneak into basket	Dwell count	3 (1.414)	4 (0.632)	F(6,5)=5.000, p=0.098	t(11)=-1.593, p=0.139	
	Hit time AOI	1040.71 (603.307)	1605.997 (1030.702)	F(6,5)=0.343, p=0.225	t(11)=-1.209, p=0.224	
	Gaze dwell time	1042.648 (588.837)	4945.413 (3197.011)	F(6,5)=0.034, p=0.001	t(5.291)=-2.948, p=0.030	Significant result
	Fix count	4.143 (1.574)	15.833 (9.923)	F(6,5)=0.025, p=0.000	t(5,213)=-2.854, p=0.034	Significant result
	TTFF AOI	1089.485 (595.523)	1636.509 (1043.462)	F(6,5)=0.324, p=0.205	t(11)=-1.185, p=0.261	
	Fixation dwell time	899.824 (357.140)	4483.446 (2918.452)	F(6,5)=0.015, p<0.000	t(5,1285)=-2.989, p=0.030	Significant result

	Mouse click count	0 (0)	0.333 (0.516)	-	*U=14, p=0.139*	
Trick question	Dwell count	3.143 (2.035)	4.167 (1.941)	-	U=12, p=0.200	
	Hit time AOI	2309.814 (2186.832)	4644.857 (5736.452)	-	U=16, p=0.534	
	Gaze dwell time	1285.456 (895.362)	8681.51 (4971.926)	F(6,5)=0.032, p=0.001	t(5.278)=-3.594, p=0.014	**Significant result**
	Fix count	4.857 (3.132)	32.333 (18.533)	F(6,5)=0.029, p=0.000	t(5.245)=-3.588, p=0.015	**Significant result**
	TTFF AOI	5699.274 (4813.076)	4650.39 (5680.218)	-	U=25, p=0.628	
	Fixation dwell time	1040.592 (577.794)	7438.561 (4631.942)	F(6,5)=0.016, p<0.000	t(5.134)=-3.361, p=0.019	**Significant result**
	Mouse click count	0 (0)	0 (0)	-	-	-
False hierarchy	Dwell count	3.288 (1.976)	4 (1.265)	F(6,5)=2.441, p=0.346	t(11)=-0.760, p=0.464	
	Hit time AOI	3898.001 (3887.047)	1927.532 (4301,485)	-	U=32, p=0.138	
	Gaze dwell time	2168.27 (1120.072)	4411.58 (1059.712)	F(6,5)=1.117, p=0.923	t(11)=-3.689, p=0.004	**Significant result**
	Fix count	7.857 (4.947)	12.5 (3.564)	F(6,5)=1.927, p=0.488	t(11)=-1.908, p=0.083	
	TTFF AOI	3922.993 (3908.914)	1958.094 (4308.56)	-	U=32, p=0.138	
	Fixation dwell time	1835.786 (751.7396)	3855.73 (1162.147)	F(6,5)=0.418, p=0.319	t(11)=-3.781, p=0.003	**Significant result**
	Mouse click count	1.143 (0.378)	1 (0)	-	*U=24, p=0.440*	
Preselection	Dwell count	2.333 (1.211)	2.857 (1.773)	F(5,6)=0.467, p=0.421	t(11)=-0.610, p=0.554	
	Hit time AOI	663.878 (403.914)	744.998 (352.457)	F(5,6)=1.313, p=0.740	t(11)=-0.387, p=0.706	
	Gaze dwell time	3215.449 (1092.521)	2054.993 (897.041)	F(5,6)=1.483, p=0.64	t(11)=2.105, p=0.059	
	Fix count	9.833 (4.167)	8.143 (3.934)	-	U=27, p=0.424	
	TTFF AOI	686.095 (396.924)	755.751 (352.597)	F(5,6)=1.267, p=0.771	t(11)=-0.335, p=0.744	
	Fixation dwell time	3099.809 (1042.5)	1828.546 (1008.153)	F(5,6)=1.069, p=0.919	t(11)=2.231, p=0.047	**Significant result**
	Mouse click count	2 (0)	1.714 (0.488)	-	U=27, p=0.21	
Low-stock/High demand	Dwell count	8.333 (5.164)	7.286 (3.729)	F(5,6)=1.918, p=0.451	t(11)=0.424, p=0.680	
	Hit time AOI	4102.365 (8652.86)	2441.21 (5953.453)	-	U=29, p=0.295	
	Gaze dwell time	10519.71 (2099.788)	9739.77 (4038.955)	F(5,6)=0.270, p=0.173	t(11)=0.425, p=0.679	

	Fix count	37.333 (3.933)	32.571 (16.791)	F(5,6)=0.055, p=0.006	t(6.759)=0.727, p=0.491	
	TTFF AOI	4130.142 (8651.043)	3108.85 (5885.272)	-	U=26, p=0.534	
	Fixation dwell time	8262.134 (2235.399)	6619.535 (3438.002)	F(5,6)=0.423, p=0.363	t(11)=0.100, p=0.339	
	Mouse click count	0.5 (0.837)	0.143 (0.378)	-	U=25.5, p=0.439	
Confirmshaming	Dwell count	1.333 (0.517)	2.429 (1.134)	-	U=8, p=0.052	.
	Hit time AOI	129.820 (164.599)	453.255 (591.299)	-	U=10, p=0.138	
	Gaze dwell time	3532.285 (746.5318)	3864.955 (1880.369)	F(5,6)=0.158, p=0.061	t(11)=-0.405, p=0.693	
	Fix count	10.667 (2.066)	13.714 (5.438)	F(5,6)=0.144, p=0.051	t(11)=-1.289, p=0.224	
	TTFF AOI	135.391 (183.135)	530.658 (608.719)	-	U=12, p=0.234	
	Fixation dwell time	3005.33 (1019.45)	2497.786 (1564.245)	-	U=30, p=0.234	
	Mouse click count	0.333 (0.516)	0.429 (0.535)	-	U=19, p=0.800	
Times	Overall duration task 1	63425.710 (21798.630)	79313.670 (25823.850)		U=30, p=0.234	
	Overall duration task 2	79113.000 (16733.480)	69572.140 (21232.850)	F(6,5)=1.610, p =0.618	t(11)=-0.888, p=0.394	
	Duration of the checkout process task 1	13320 (8267.91)	23450.5 (7395.291)	F(6,5)=1.250, p=0.825	t(11)=-2.310, p=0.041	**Significant result**
	Duration of closing cookie pop-up	5151.833 (1101.709)	6629.714 (6738.036)	-	U=26, p=0.534	
	Time to close the confirmshaming popup (green one)	3877.667 (484.2609)	4854.714 (2083.501)	F(6,5)=18.511, p=0.006	t(6.748)=-1.203, p=0.269	

Table 10 Overview of statistical analysis.