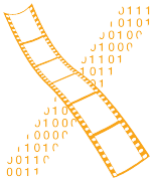


Chair for Computer Science 10 (Media Computing and Human-Computer Interaction)

RWTH AACHEN UNIVERSITY

Designing Countermeasures Against Dark Patterns in User Interfaces

Master's Thesis
submitted to the
Media Computing Group
Prof. Dr. Jan Borchers
Computer Science Department
RWTH Aachen University

by
Paul Miles Preuschoff

Thesis advisor:
Prof. Dr. Jan Borchers

Second examiner:
Prof. Dr. Ulrik Schroeder

Registration date: 28.02.2023
Submission date: 22.08.2023

Zentrales Prüfungsamt/Central Examination Office

Eidesstattliche Versicherung
Statutory Declaration in Lieu of an Oath

Name, Vorname/Last Name, First Name

Matrikelnummer (freiwillige Angabe)
Matriculation No. (optional)

Ich versichere hiermit an Eides Statt, dass ich die vorliegende Arbeit/Bachelorarbeit/ Masterarbeit* mit dem Titel
I hereby declare in lieu of an oath that I have completed the present paper/Bachelor thesis/Master thesis* entitled

selbstständig und ohne unzulässige fremde Hilfe (insbes. akademisches Ghostwriting) erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt. Für den Fall, dass die Arbeit zusätzlich auf einem Datenträger eingereicht wird, erkläre ich, dass die schriftliche und die elektronische Form vollständig übereinstimmen. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.
independently and without illegitimate assistance from third parties (such as academic ghostwriters). I have used no other than the specified sources and aids. In case that the thesis is additionally submitted in an electronic format, I declare that the written and electronic versions are fully identical. The thesis has not been submitted to any examination body in this, or similar, form.

Ort, Datum/City, Date

Unterschrift/Signature

*Nichtzutreffendes bitte streichen

*Please delete as appropriate

Belehrung:
Official Notification:

§ 156 StGB: Falsche Versicherung an Eides Statt
Wer vor einer zur Abnahme einer Versicherung an Eides Statt zuständigen Behörde eine solche Versicherung falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.
Para. 156 StGB (German Criminal Code): False Statutory Declarations
Whoever before a public authority competent to administer statutory declarations falsely makes such a declaration or falsely testifies while referring to such a declaration shall be liable to imprisonment not exceeding three years or a fine.
§ 161 StGB: Fahrlässiger Falscheid; fahrlässige falsche Versicherung an Eides Statt
(1) Wenn eine der in den §§ 154 bis 156 bezeichneten Handlungen aus Fahrlässigkeit begangen worden ist, so tritt Freiheitsstrafe bis zu einem Jahr oder Geldstrafe ein.
(2) Straflosigkeit tritt ein, wenn der Täter die falsche Angabe rechtzeitig berichtigt. Die Vorschriften des § 158 Abs. 2 und 3 gelten entsprechend.
Para. 161 StGB (German Criminal Code): False Statutory Declarations Due to Negligence
(1) If a person commits one of the offences listed in sections 154 through 156 negligently the penalty shall be imprisonment not exceeding one year or a fine.
(2) The offender shall be exempt from liability if he or she corrects their false testimony in time. The provisions of section 158 (2) and (3) shall apply accordingly.

Die vorstehende Belehrung habe ich zur Kenntnis genommen:
I have read and understood the above official notification:

Ort, Datum/City, Date

Unterschrift/Signature

Contents

List of Figures

List of Tables

Abstract

As the Internet becomes an integral part of modern life, users are increasingly exposed to online manipulation. Dark Patterns are a common form of online manipulation used by online service providers to maximize their financial gain by manipulating consumer decisions. Dark Patterns are elements on websites that use interface design principles and exploit cognitive biases to influence users against their own interests. These practices pose risks such as financial loss, invasion of privacy, and wasted time. To date, research has focused primarily on categorizing Dark Patterns and their negative impact on users. Some work has been done to automatically detect Dark Patterns. Detection alone does not help users, but it can be used to mitigate negative effects, make users more resilient, or put public pressure on companies. In our work, we investigate the question *How could visual responses of Dark Pattern detecting systems improve browsing the web?* We quantitatively and qualitatively compared different visual countermeasures in two studies. Study 1 was an online study with 40 participants. It provided a first broad understanding of user perception. In Study 2, we narrowed down the set of countermeasures and conducted an interactive lab study with 20 participants. We found that users generally prefer countermeasures that are information-rich and transparent. In specified situations, user preferences are mostly heterogeneous. Our countermeasures can improve the usability of a website, slow users down, and affect how much users trust a website. Our work provides valuable insights from a human-computer interaction perspective to inform the design of systems that mitigate the negative effects of Dark Patterns.

Überblick

Da das Internet zu einem integralen Bestandteil des modernen Lebens geworden ist, sind Nutzer zunehmend Online-Manipulation ausgesetzt. Dark Patterns sind eine gängige Form der Online-Manipulationen, die von Dienstanbietern genutzt wird, um den finanziellen Gewinn durch die Manipulation von Verbraucherentscheidungen zu maximieren. Dark Patterns sind Elemente auf Webseiten, die Prinzipien des Nutzeroberflächendesigns anwenden und kognitive Eigenschaften ausnutzen, um Nutzer gegen ihre eigenen Interessen zu beeinflussen. Diese Praktiken bergen Risiken wie finanziellen Verlust, Verletzung der Privatsphäre und Zeitverschwendung. Bisher haben sich Forscher vor allem auf die Kategorisierung von Dark Patterns und deren negative Auswirkungen auf die Nutzer konzentriert. Es wird ebenfalls erforscht Dark Patterns automatisch zu erkennen. Methoden zur automatischen Erkennenung könnten genutzt werden, um die negativen Auswirkungen von Dark Patterns zu mildern, die Nutzer widerstandsfähiger zu machen oder öffentlichen Druck auf Unternehmen auszuüben. In dieser Arbeit gehen wir der Frage nach *Wie könnten visuelle Reaktionen von Systemen zur Erkennung von Dark Patterns das Surfen im Internet verbessern?* In zwei Studien haben wir verschiedene visuelle Gegenmaßnahmen quantitativ und qualitativ verglichen. Studie 1 war eine Online-Studie mit 40 Teilnehmern. Sie lieferte ein erstes breites Verständnis der Nutzerwahrnehmung. In Studie 2 haben wir unsere Auswahl an Gegenmaßnahmen weiter eingegrenzt und eine interaktive Laborstudie mit 20 Teilnehmern durchgeführt. Wir fanden heraus, dass die Nutzer im Allgemeinen Gegenmaßnahmen bevorzugen, die informationsreich und transparent sind. Für vorgegebene Situationen sind die Präferenzen der Nutzer meist heterogen. Unsere gegenmaßnahmen können die Nutzerfreundlichkeit einer Website verbessern, die Nutzer verlangsamen und das Vertrauen der Nutzer in eine Website beeinflussen. Unsere Arbeit liefert wertvolle Erkenntnisse aus der Perspektive der Mensch-Computer-Interaktion für die Entwicklung von Systemen, welche die negativen Auswirkungen von Dark Patterns abmildern.

Acknowledgements

I would like to thank Prof. Dr. Jan Borchers and Prof. Dr. Ulrik Schroeder for reviewing my thesis.

I would also like to thank the people at i10 and the Council of Renés for giving me the opportunity to be part of this interesting project. In particular, I would like to thank them for inviting me into the world of real, breathing science. Thank you, René Schäfer and Dr. René Röpke, for your constructive feedback, helpful expertise, and engaging motivation on this topic!

A big thank you to all the people who participated in the studies. You took the time and effort to make this work possible.

Thank you, Inken Kramme, for agreeing to meet and talk about the Dark Pattern Detection Project.

Finally, I would like to thank my family, friends and girlfriend. Thank you for your advice and embracing support!

Conventions

Throughout this thesis we use the following conventions.

Definitions of technical terms are set off in colored boxes.

EXAMPLES:
An example is something that is typical for the group it belongs to. It can be used as a way to help someone understand a concept.

Definition:
Examples

For better clarity, numbers are written as figures when referring to quantities: For example: 8 participants liked this feature.

This thesis is written in American English.

The first person is written in the plural form. Unidentified third persons are referred to in the plural form.

Chapter 1

Introduction

The Internet and its numerous online services have become an integral part of everyday life. Thus, the economic interests of service providers such as streaming services or online shops make online manipulation an increasingly relevant phenomenon that affects many. Online manipulation can be used, for example, to influence shopping behavior [Mathur et al., 2019, Voigt et al., 2021] or privacy decisions [Utz et al., 2019, Graßl et al., 2021] of users in favor of the service provider and against their own interests. Such practices deliberately target the vulnerabilities of individuals outside their conscious awareness to decrease their ability to decide independently [Susser et al., 2019]. Other, more aggressive methods might be noticed by users and annoy, frustrate or anger them [Gray et al., 2020, Maier and Harr, 2020].

Online manipulation negatively affects everyday life

1.1 Dark Patterns

DARK PATTERN (DP):
According to Mathur et al. [2019] Dark Patterns (DPs) are "user interface design choices that benefit an online service by coercing, steering, or deceiving users into making decisions that, if fully informed and capable of selecting alternatives, they might not make".

Definition:
Dark Pattern (DP)

DPs are online manipulation through design and phrasing

DPs are a specific form of online manipulation that uses interface design principles and exploits cognitive biases to influence users against their own interest [Mathur et al., 2019]. They are widely applied and can have negative impacts on individual welfare. For example, they can invade privacy and cause financial loss or behavioral addictions [Mathur et al., 2021]. Even users that are generally aware of the concept of DPs stay susceptible to them [Bongard-Blanchy et al., 2021].

The field of DP research is barely more than a decade old

The term *Dark Pattern* was established by the UX researcher Harry Brignull in 2010. Since then, researchers have started searching and categorizing different DP types that were applied in practice [Gunawan et al., 2021, Mathur et al., 2019, Di Geronimo et al., 2020a, Gray et al., 2018]. Depending on perspective and subject, different DP taxonomies have evolved in the last decade [Mathur et al., 2021].

Cookie consent requests are known to make heavy use of DPs

Cookie consent requests are a popular application area for DPs [Graßl et al., 2021]. They were introduced to websites to manage privacy settings for cookies on a page. Service providers have an interest in processing and selling user data, while users have no advantage from that. By implementing DPs, service providers can manipulate users towards choices that prioritize privacy less [Nouwens et al., 2020].

1.2 Visual Countermeasures

Visual DP countermeasures require an effective combination of detection and reaction methods

Since general awareness does not make users resistant against DPs [Bongard-Blanchy et al., 2021], it seems necessary to develop systems that introduce a counter force against DPs. So far, research regarding this subject focuses primarily on the automatic detection of DPs [Mathur et al., 2019, Soe et al., 2022], but not on reacting to detections. Assuming that reliable detection mechanisms for DPs are developed, their results could be used to adjust the user interface to potentially be less manipulative. We call these adjustments *visual countermeasures* . Building on suggestions from other researchers [Mathur et al., 2019, Conti and Sobiesk, 2010, Graßl et al., 2021, Moser et al., 2019, Bongard-

Blanchy et al., 2021], we investigate different methods to visually alter websites to mitigate negative DP effects.

1.3 Outline

The scope of this thesis is to look ahead and investigate how visual DP countermeasures (for example in the form of a browser add-on) should behave when a DP is detected. What is the user's perception of different visual countermeasures? How do different visual countermeasures impact the interaction with pages that use multiple DPs? In this thesis, we address those questions to provide a scientific foundation for developers and researchers to create anti-DP systems that suit user needs.

This thesis focuses on the design of visual DP countermeasures

We start by sketching out the research area of DPs by explaining what they are, why they are a problem, how they can be classified and how they could be counteracted (Chapter 2 "Related Work"). We continue by briefly describing in Chapter 3 "Considerations on Dark Pattern Countermeasures" the general approaches against DPs that we identified, based on related work. Additionally, we sketch out a set of ideas for visual countermeasures that are based on literature and that we developed ourselves.

We first present related work and our own ideas

From this set of ideas, we selected a subset of seven comparable visual countermeasures (including one baseline) which we then investigated in a screenshot-based online study (Chapter 4 "Study 1: Online Study"). This first study provided a basic understanding of the user perspective. We present our results and discuss them. Next, we closer investigated four of these methods (including one baseline) with an interactive lab study, which allowed us to observe a more realistic setting with real interaction and a larger set of DPs (Chapter 5 "Study 2: Lab Study"). We discuss the results from Study 2 in comparison with the results from Study 1 in Chapter 6 "Evaluation". Finally, we summarize our findings and sketch out promising avenues for future work in Chapter 7 "Summary and Future Work".

We conducted an online study for basic understanding and a lab study for deeper insights

Chapter 2

Related Work

In this section, we provide an overview of the field of DPs and why they are an important contemporary problem. We then present the taxonomy we use in this thesis to categorize DPs. Next, we outline the general ways in which society can counteract DPs. Based on this context, we discuss the current state of research on DP detection and how human-computer interaction research can contribute to the development of visual DP countermeasures.

We present related work on DPs and how to detect and counter them

2.1 Dark Patterns and their Implications

In 2010, Conti and Sobiesk [2010] pointed out the emergence of malicious interface design , and discussed its impact on users. In the same year, the term *Dark Pattern* was established by the UX researcher Harry Brignull on his website[1], dedicated to collect and discuss examples for manipulative user interface designs.

DP research dates back to 2010

DPs can be classified as a subfield of malicious interface design. In the research community, many attempts to define DPs have been made [Mathur et al., 2021]. Common themes of these definitions are that DPs are intentionally

It is often not clear if something is a DP, definitions vary

[1] `https://www.deceptive.design/` (former adress: `https://www.darkpatterns.org`) accessed April 2023

designed to subvert user preferences or intent to benefit the provider of a service [Mathur et al., 2021]. It is not always easy to draw the line between DPs and clever marketing. An example for this are online shops that target the users' impulse buying behavior by using compelling, high quality product photos or videos [Moser et al., 2019]. This might appear to be a DP at first glance, but we do not classify it as such because this method does not deceive.

In the following, we will discuss that DPs are a threat for two main reasons: they are effective tools of manipulation, and they are widely used.

DPs have a number of negative effects on users

This can have a huge negative effect on users. Luguri and Strahilevitz [2021] showed that DPs can make it two to four times more likely for customers to subscribe to dubious services, depending on how strongly DP principles are applied. Especially DPs like *Hidden Information*, *Trick Questions*[2] and obstruction strategies had a notable impact. While the use of DPs might damage the reputation of a brand or the users' trust in the service [Maier and Harr, 2020, Bhoot et al., 2021, Voigt et al., 2021], it seems like a less obvious use of DPs can have a big impact on users without alienating them [Luguri and Strahilevitz, 2021]. This can result in negative effects for users, including economic distress, frustration, increased mental and physical effort, wasting time, and invasion of privacy [Conti and Sobiesk, 2010, Mathur et al., 2019, Baroni et al., 2021, Bhoot et al., 2021, Bösch et al., 2016].

DPs are almost ubiquitous on the modern web

Additionally, it is hard to escape DPs, since they are widely applied. Mathur et al. [2019] found DPs on 11.1% of the 11k shopping pages they examined. The authors stress that this is only a lower bound, the actual percentage is expected to be drastically higher. DPs seem to be especially common on more popular websites [Mathur et al., 2019]. Di Geronimo et al. [2020b] found DPs in 95% of the 240 popular mobile apps they analyzed. There is often no equal alternative to large service providers that use DPs, so even if the users get frustrated, DPs are almost impossible to get around [Maier and Harr, 2020, Bhoot et al., 2021]. Surpris-

[2]For more details and explanations of the different DPs, see Mathur et al. [2019]

ingly, this frustration might even result in a reduced willingness to actively counteract the DPs, further increasing their impact [Maier and Harr, 2020]. In general, many users seem to understand the existence of DPs, but do not understand how they are personally affected by them [Bongard-Blanchy et al., 2021]. DPs are viewed by many as evil, resulting in phenomena like the *r/assholedesign* Subreddit[3] [Gray et al., 2020]. In this forum, instances of malicious design, including DPs, are shared and discussed.

2.1.1 Types of DPs

A variety of attempts were made to categorize DPs [Mathur et al., 2019, Hogan et al., 2022, Gunawan et al., 2021, Gray et al., 2020, Moser et al., 2019, Gray et al., 2018, Conti and Sobiesk, 2010, Bhoot et al., 2021]. A recent overview over the different DP taxonomies is provided by Mathur et al. [2021]. Depending on perspective and research setting, different taxonomies are preferable. For example, DPs vary strongly between desktop, mobile and app [Gunawan et al., 2021]. For this thesis, we decided to reference the taxonomy introduced by [Mathur et al., 2019], since it focuses on shopping websites for desktop and fits thus well the setting of our research. The taxonomy consists of 15 DPs organized in seven categories (Tab. 2.1).

We chose the DP taxonomy from Mathur et al. [2019] because it is based on shopping websites

2.1.2 Leverages against DPs

It seems necessary to counteract the described negative impacts of DPs. This can be addressed in multiple different ways. Gray et al. [2018] suggest to increase the awareness of UX practitioners on the moral implications of DP implementation. However, it seems unlikely that this alone will be a solution, since it does not address the root of the problem. It might be more effective to either focus on reducing the impact of DPs on users or getting the companies less

To address DPs, one can weaken the DPs or strengthen the users' ability to handle them

[3] `https://www.reddit.com/r/assholedesign/` (accessed april 2023)

Category	Type
Sneaking	Sneak into Basket
	Hidden Cost
	Hidden Subscription
Urgency	Countdown Timer
	Limited-time Message
Misdirection	Confirmshaming
	Visual Interference
	Trick Question
	Pressured Selling
Social Proof	Activity Message
	Testimonials
Scarcity	Low-stock Message
	High-demand Message
Obstruction	Hard to Cancel
Forced Action	Forced Enrollment

Table 2.1: This excerpt from the DP taxonomy by Mathur et al. [2019] will be used throughout this thesis.

inclined to implement DPs [Graßl et al., 2021]. Bongard-Blanchy et al. [2021] introduced an intervention space that is primarily focused on the user. They suggest that scopes of interventions could be to raise awareness of users, facilitate DP detection, increase resistance towards DPs or eliminate DP from online services. It is additionally frequently suggested to focus on the companies that implement DPs in their services by taking legislative measures [Graßl et al., 2021, Bongard-Blanchy et al., 2021, Nouwens et al., 2020]. However, it can also happen that legal constraints cause DPs, rather than prevent them [Chromik et al., 2019]. This happened for example when the EU's General Data Protection Regulation[4] made consent management platforms for cookies necessary [Nouwens et al., 2020]. Service providers were forced to let users decide whether or not to accept cookies for tracking and targeted advertising. Since users had no incentive to accept these cookies, DPs were implemented to trick users into accepting them anyway.

[4] `https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679` (accessed: April 2023)

Furthermore, increasing the awareness of users seems to be surprisingly ineffective against DPs. Studies by Bongard-Blanchy et al. [2021] showed that users who easily *recognize* manipulative elements online have a lower likelihood of being manipulated by them. However, their *awareness* of online manipulation does not influence their average likelihood of being influenced. Additionally, Voigt et al. [2021] conducted a study in which they found no correlation between participants' familiarity with technology and their ability to recognize DPs. It seems like the sole knowledge of DPs existence does not protect users against deception [Maier and Harr, 2020]. Making DPs easier to recognize might help [Bongard-Blanchy et al., 2021].

Raising DP awareness is not enough, technical support might help

2.2 Visual Countermeasures

It is already possible to detect some DPs automatically , for example, by using web crawlers [Mathur et al., 2019], pattern matching[5] or machine learning [Soe et al., 2022]. However, DP detection technology is currently under development and not mature yet. Certain types of DPs might even be impossible to detect automatically [Curley et al., 2021]. In those cases, community-driven approaches could be used to identify them [Kollnig et al., 2021].

DP detection is a topic of current research

Given that DPs can be detected and classified reliably, different approaches were suggested in related work on how to represent this information in a user interface. Some researchers suggest a browser add-on, comparable to add blockers [Mathur et al., 2019, Conti and Sobiesk, 2010]. Programs like this could generate warnings or explanations [Graßl et al., 2021, Mathur et al., 2019][6] or they could remove the manipulative elements [Moser et al., 2019]. Bongard-Blanchy et al. [2021] pointed out that warnings might be problematic in this context, as they might become

Detecting a DP is not enough, the system needs to respond in a helpful way

[5] We talked to researchers from the German DAPDE (Dark Pattern Detection) project (`https://dapde.de`), they told us this is that pattern matching is their current approach

[6] As part of a hackathon the work of Mathur et al. [2019] was used in such a way for a Chrome extension `https://github.com/NicholasTung/dark-patterns-recognition`

rapidly ineffective due to warning fatigue.

This thesis focuses on the visual reaction to DP detection

From the literature review, we conclude that visually countering DPs with add-on like systems is a promising approach. To our knowledge, the best way to communicate the detection of a DP to the user has not been explored. This important next step in effectively counteracting DPs is the subject of this thesis.

Chapter 3

Considerations on Dark Pattern Countermeasures

This chapter details the considerations that guided us in defining the exact scope of our research. Because the field of DP countermeasures is still in its infancy, we begin by describing in general terms why DPs exist at their current severity. By understanding the forces that drive service providers to use DPs, we can gain a better understanding of how to reduce them. This results in four different leverage points for countering DPs. We next present countermeasure ideas that use these leverages to varying degrees. We identify characteristics that some of our ideas have in common and select a subset that is suitable for our research. Finally, we define our underlying assumptions about the technical context, since all of our prototypes for studies are built manually in a Wizard of Oz-like fashion.

We provide general context for our scope of research

3.1 Leverages against DPs

This is our working model of why DPs are used in their current form. Companies use DPs to maximize their financial gain [Maier and Harr, 2020]. This is the force that increases

Figure 3.1: The forces that increase or decrease the use of DPs by service providers. To reduce DP usage, the force that increases it can be reduced, or an additional force that decreases it can be introduced.

the use of DPs on the Web. For example, an online store might include a fake countdown for a sale, which increases the amount of products sold. Two other forces oppose the use of DPs: annoyed users and legal restrictions (Fig. 3.1). With this model in mind, DP countermeasures can either strengthen the forces already working against DPs, introduce additional forces working against DPs, or reduce the force driving DP use.

DPs are implemented to maximise financial gain

Based on these considerations and on the related work, especially by Gray et al. [2021], four possible points of leverage against DPs can be identified:

DPs can be counteracted by laws and public pressure; their impact can be reduced by awareness and technical systems

- **Laws**: Additional laws could strengthen the counter force against DP use. By banning DPs and enforcing these restrictions, DPs could theoretically be permanently removed from the Web. However, it is very difficult and resource intensive to enforce such restrictions [Luguri and Strahilevitz, 2021].
- **Public Pressure**: By revealing the manipulative practices of certain service providers, public pressure on them would be increased, making DPs financially unfeasible for them. This could remove DPs from the Web to some extent, but it would have to be realized in a fair and transparent way.
- **Awareness of Users**: By making users more aware of the concept of DPs and how they are manipulated, users might become less susceptible to their influence. This would reduce the presence of DPs in the long run by making them less effective for service providers. However, this approach requires more mental effort

on the part of users and cannot make users immune to DPs [Bongard-Blanchy et al., 2021].

- **Technically weaken DPs**: Software on the users' device that automatically changes or removes DPs would not require large scale changes. It could mitigate the negative effects of the DPs. However, building such a tool is technically challenging and, as we will show in this thesis, not all DPs can be removed on the client side.

In general, these four dimensions of leverage can be combined and used by countermeasures against DPs. Since the first proposed dimension requires a juristic approach, in this thesis, we will focus on the other three dimensions. They allow us to organize ideas for visual countermeasures.

We will consider three dimensions of leverage against DPs

3.2 General Ideas for Countermeasures

After reviewing the literature, we conducted a brainstorming session with four researchers. The goal was to find approaches for visual countermeasures that could be implemented in a browser add-on or similar systems. The most interesting ideas are sketched out below. Ideas that we investigate in our studies are marked with an $*$.

A brainstorming session resulted in 11 intriguing ideas for technical countermeasures

- **General Tips**: By providing educative tips similar to loading screen tips in computer games to users, their awareness could be increased. However, the tips would most likely be ignored and lack relevant context.
- **Highlighting**$*$: By clearly marking DPs on a page, users would be warned and could be more careful. However, they would not learn much. This approach requires automatic DP detection.
- **Highlighting with Explanation**$*$: Combining *Highlighting* with the *General Tips*, results in a method that

provides tips with context and warns the user. However, it might result in too much visual clutter on a page. Automatic DP detection and classification is required. This method was suggested by Mathur et al. [2019].

- **Lowlighting**$*$: Similar to *Highlighting*, but instead of visually directing focus towards the DP, it is recolored to be less alarming. It could also be well combined with explanations.
- **Cursor Highlighting**: The page looks the same, but if the user hovers over an area that contains a DP the cursor indicates the presence of a DP. This is less intrusive than *Highlighting*, but only works on DPs that are hovered over. This approach requires DP detection.
- **Hiding**$*$: DPs are removed or replaced with non-manipulative versions. The user is ideally not exposed to manipulation, but has to trust the countermeasure as it changes page content. This approach requires DP detection and removal.
- **Hiding with Marking**$*$: Similar to *Hiding*, but areas that were changed get marked. This provides additional information to the user.
- **Switch**$*$: Similar to *Hiding*, but areas that were changed can be toggled individually to the original state.
- **Page Score**: Introduce an indicator for the amount of manipulation happening on a page, similar to the Nutri-score on food[1]. This approach requires DP detection and a heuristic to compute the score of a page.
- **Page Score in Search Engine**: Uses the *Page Score*, but instead of showing the score when the user is already on a page, it could help users to choose a website if the score would already be shown in the list of possible websites.

[1] `https://www.bmel.de/DE/themen/ernaehrung/lebensmittel-kennzeichnung/freiwillige-angaben-und-label/nutri-score/nutri-score_node.html` (accessed March 2023)

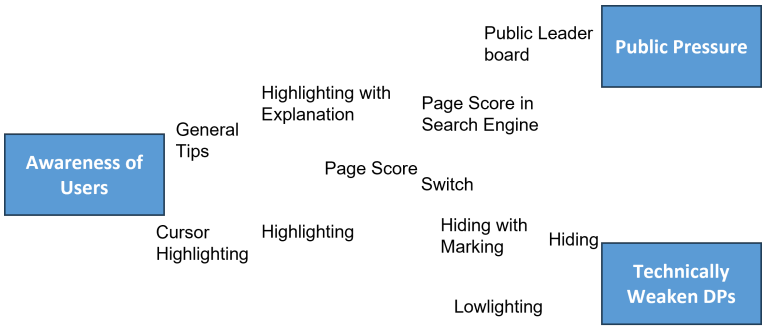


Figure 3.2: Our 11 ideas for potential technical DP countermeasures organized based on three dimensions of leverage that they use. Higher proximity to a dimension means that this point of leverage is used heavily.

- **Public Leaderboards**: By publishing lists of popular websites that use DPs heavily, the public pressure might result in the service providers reducing the DPs.

These methods are organized in Figure 3.2 based on the three dimensions of DP leverage described above (leaving out *Laws*).

3.3 Design Choices

The ideas above can be combined in different ways and share different sets of characteristics. We identified some choices when designing DP countermeasures:
Should the countermeasure...

DP countermeasures can have different characteristics

- act on the page itself or somewhere else?
- act globally on the page, or for each element individually?
- alter existing content of a page or only add content?
- provide additional information or be discrete?
- focus attention on the DP or divert attention from it?

3.4 Countermeasure Behavior Definitions

We have made assumptions to provide a realistic foundation for our research

In this section, we define how the countermeasures behave in our prototypes. To keep them grounded on the same realistic foundation, we make the following assumptions:

- The system is able to reliably detect and classify DPs automatically.
- Countermeasures can alter the visual appearance of manipulative elements. They can not change any functionality of a page.
- Content that is revealed only later in the interaction process is unknown to the system until it occurs.
- The system does not make autonomous interaction decisions (like automatically rejecting cookies).

In this thesis, we focus on six visual countermeasures

Regarding the design choices mentioned above, we will focus on visual countermeasures that are located on the page itself and that are individual for each element. We exclude *Cursor Highlighting*, because it can not be represented with our prototyping tools. This way, we obtained a subset of six visual countermeasures that have a comparable context. Implementing the countermeasures in our studies requires a precise definition of their appearance and behavior to keep them consistent. A symbolic representation of them is provided in Figure 3.3. They are defined as follows:

- *Unchanged* (*UC*): Manipulative elements are not changed in any way. *UC* serves as the baseline.
- *Highlight with Explanation* (*HL+E*): A box with a red dashed line is drawn around the areas of manipulative elements. Next to each box, a small red warning sign is added. It includes a tooltip that explains why content is marked. The explanation outlines what exactly is manipulative and how it aims to affect the user. It is based on the type of DP that was used. A similar idea was proposed by Mathur et al. [2019] as future work.

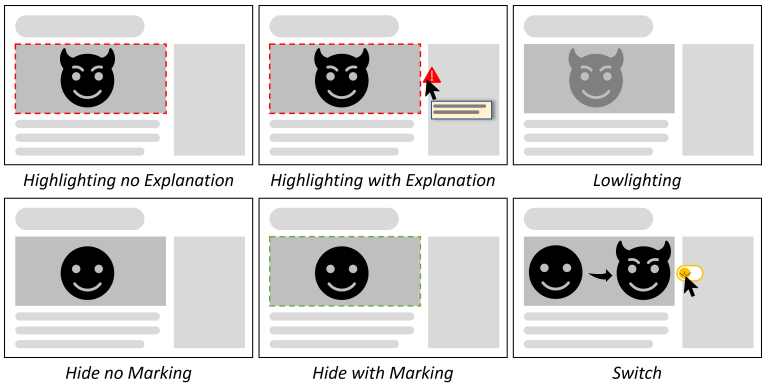


Figure 3.3: Symbolic representations of the six countermeasures examined in this thesis. The evil face represents a DP, the friendly face represents a neutral version of the element. Refer to Section 3.3 for detailed explanations.

- *Highlight without Explanation* (*HL*): Similar to *HL+E*, but without the warning signs and explanations. This might reduce distraction during normal interaction with the page.
- *Lowlight* (*LL*): Manipulative elements are adjusted to be less visible, for example by graying out texts or buttons. This strategy tries to deviate attention away from the element, ideally weakening its impact on the user.
- *Hide without Marking* (*HD*): Manipulative elements are removed or replaced by a non-manipulative version. This can include rephrasing of text or labels. It can not be applied to all DP types.
- *Hide with Marking* (*HD+M*): Similar to *HD*, but a box with a green dashed line is included around the areas of manipulative elements to indicate the change. This way, users should be able to tell what is original page content, and where the countermeasure is active.
- *Switch* (*SW*): Similar to *HD*, but next to changed areas an icon is shown. When clicked, the icon changes to indicate deactivation and the manipulative element is shown instead. This allows users to access the original content.

Chapter 4

Study 1: Online Study

In this chapter, we describe the design of Study 1, present the results, and discuss them. The purpose of this study is to gain basic insight into user perceptions of the various countermeasures. It is an online study that uses screenshot prototypes and collects rankings, ratings, and qualitative data.

4.1 Study Procedure

Study 1 compared six countermeasures on three DPs

The setting for Study 1 was a single view of a fictional online shop for sports shoes. It was adapted from Bongard-Blanchy et al. [2021]. The scenario was shown to participants as different variations of screenshots[1]. Each screenshot included one of three common DPs: *Confirmshaming*, *Low Stock Message*, and *Visual Interference* (Fig. 4.1). We decided to pick these patterns because:

- They are among the most frequently used ones according to Mathur et al. [2019].
- They can be represented on a single view (unlike *Hidden Cost* for example).

[1] created in PowerPoint `https://www.microsoft.com/en-en/microsoft-365/powerpoint?rtc=1` (accessed April 2023)

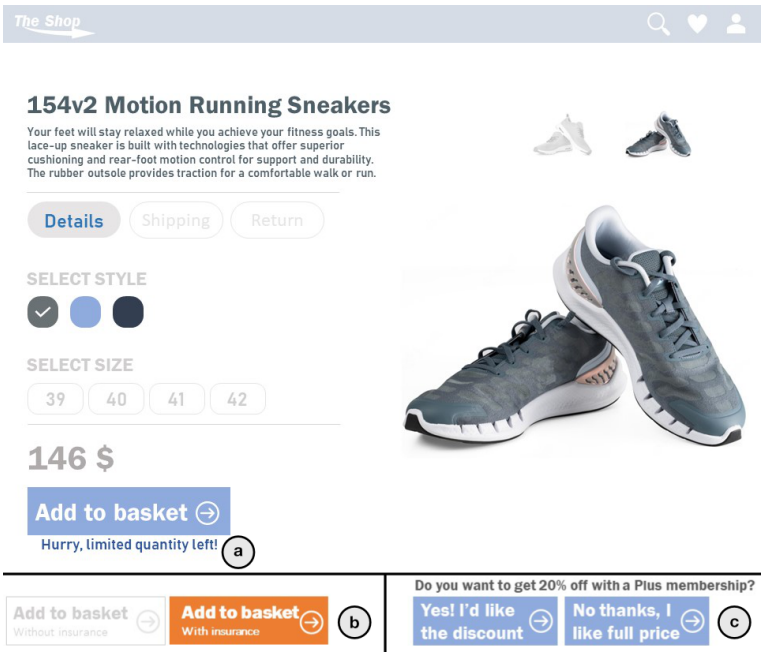


Figure 4.1: Overview of the three DPs used in Study 1: (a) *Low Stock Message*, (b) *Visual Interference*, (c) *Confirmshaming*. In each screenshot, only one pattern was present. We applied all DPs to the "Add to basket" button for consistency.

- The DPs use wording or visual design in different ways, we expect that different strength and weaknesses of the countermeasures become visible.

In Study 1, we compared all six countermeasures and the baseline defined in Section 3.3. The combinations of DPs and countermeasures are presented in Figure 4.2.

4.1.1 Part 1: Rankings for Different DPs

Participants ranked the countermeasures in specific situations and in general

The study consisted of two parts. The goal of the first part was to find out which countermeasure the participants preferred for each DP. For this, we showed screenshots of a DP being counteracted by the different countermeasures. The screenshots are provided in Appendix A. Participants were

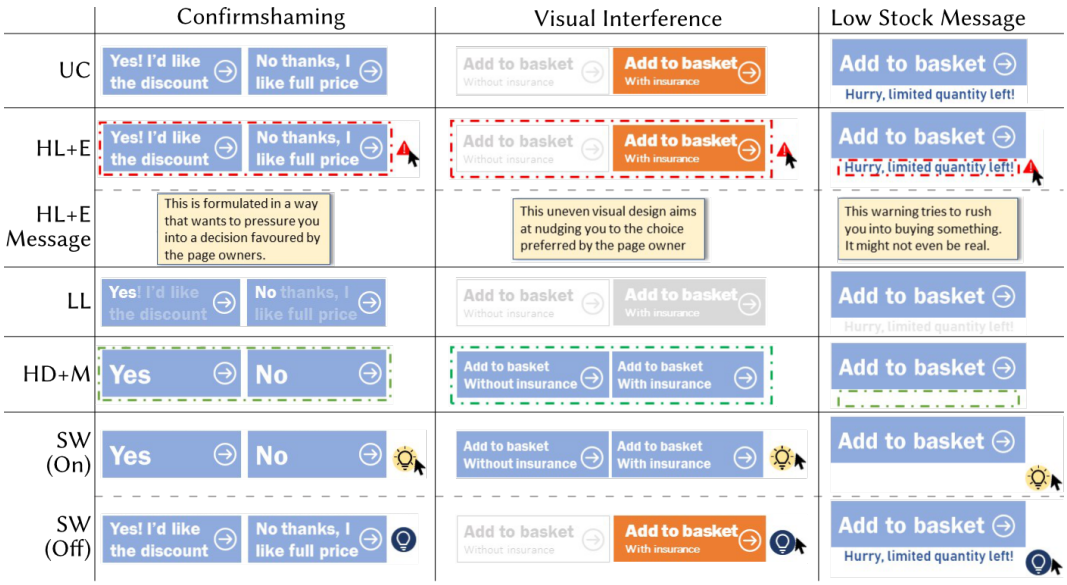


Figure 4.2: Overview over the countermeasures (Sec. 3.3). Not all countermeasures are directly depicted in this figure: *HL* looks like *HL+E* without the red warning sign and explanations. *HD* is similar to *HD+M*, but does not include the green box.

asked to rank the countermeasures for the given DP. This was repeated for each of the three DPs. We included an overall ranking at the very end of the study. We decided to put it last in order to maximize the time participants could reflect on the countermeasures.

4.1.2 Part 2: Ratings and Comments

The second part of the study focused on general opinions regarding each individual countermeasure. At this point, the countermeasures were already seen in action in the first part. Participants were asked to fill out six 7-point semantic differential ratings regarding *efficiency*, *easiness*, *clarity*, *helpfulness*, *safety*, and whether a website would *feel better or worse* with the countermeasure. To qualitatively gather the opinions of participants, the study also included questions on the strengths and weaknesses of each countermeasure.

We collected participants' opinions using ratings and free text

4.1.3 Study Design

Study 1 used a within-subject design

The study was designed to take around 30 minutes. We used a within-subjects design, meaning that all participants saw all DPs and countermeasures. This was particularly helpful in collecting meaningful rankings of the countermeasures. The order of the rankings in the first part and of the rating tasks in the second part was randomized. The overall ranking remained at the end of the survey. The order of the screenshots was kept constant between rankings to reduce potential confusion of the different countermeasures. The study was created and conducted using the online survey tool *SoSci Survey*[2]. We invited our participants over private and professional connections and made public posts on Reddit in the Subreddit *r/SampleSize*[3].

4.2 Results

This section presents the quantitative and qualitative results we obtained from Study 1.

4.2.1 Participants

40 participants, mostly young adults with a technical background

Overall, 40 people participated in our study, age 20 to 62 (M = 30.45 years, SD = 11.49 years, 12 women, 25 men, 2 non-binary, and 1 n/a). 32 of them reported professions we classified as a technical academic background. They reported an average daily active computer browsing time of 5.58 hours (SD = 3.24, min = 0, max = 12). During this time, they reported to mostly learn or work (36) or read articles (33) (Tab. 4.1).

[2] `https://www.soscisurvey.de/` (accessed April 2023)

[3] `https://www.reddit.com/r/SampleSize/` (accessed April 2023)

Activity Category	#Participants
Working / Learning	36
Reading Articles / Blogs	33
Personal Communication	24
Shopping / Comparing Products	23
Playing Games / Entertainment	21
Social Networks	21
Online Banking and Investments	19
Other	3

Table 4.1: What participants reported doing during their active browsing time on a computer. These options were provided in advance. Participants had the option for additional free answers, but barely used it. (n = 40)

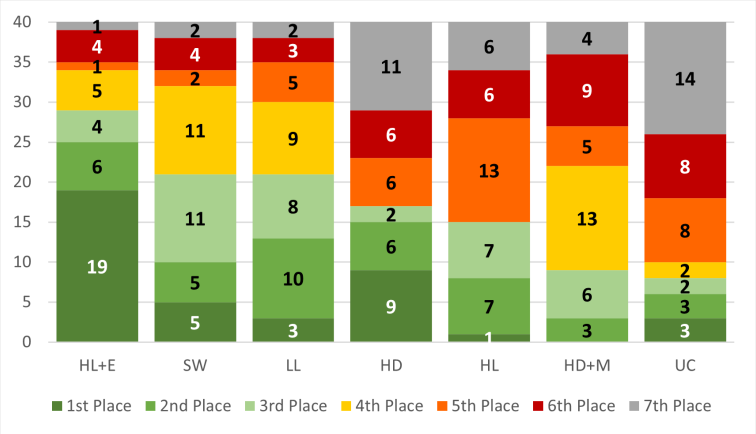


Figure 4.3: The overall rankings of the countermeasures, ordered from best to worst average ranking. *Highlight with Explanation* (*HL+E*) was ranked the best, *Hide without Marking* (*HD*) received many best and worst placements, while *Unchanged* (*UC*) received the lowest ranking. (n = 40)

4.2.2 Ranking

In the overall ranking (Fig. 4.3), *Highlight with Explanation* (*HL+E*) was ranked very highly, with 19 participants giving it the best rank. *Switch* (*SW*) and *Lowlight* (*LL*) were also rather liked by participants, while *Highlight without Expla-*

nation (*HL*) and *Hide with Marking* (*HD+M*) were rather disliked. The baseline *Unchanged* (*UC*) was ranked last by 14 of our 40 participants, but was also rated best 3 times. *HD* was by far most controversial: It was ranked best 9 times, but also ranked worst 11 times. The DP specific rankings (Fig. 4.4) differed from the general rankings. When comparing the median of the rankings (Fig. 4.5), *HD* changed drastically. For *Visual Interference* it was ranked good and for *Low Stock Message* it was ranked very good compared to the general ranking. *SW* was in general ranked well, but the median dropped by one place for *Low Stock Message* compared to other ranking contexts. Similarly, the median rank of *HL+E* rose by one place for *Confirmshaming* in comparison. While *LL* reacted strongly on the ranking context, *UC* and *HL+E* stayed rather consistent.

In general *HL+E* was ranked best, *HD* was ranked most controversial

HD was ranked good in some specific contexts

4.2.3 Ratings

We asked participants to rate each countermeasure (except the baseline) in six categories using 7-point semantic differential scales. This means that we provided for each category a scale with seven steps, with a negative attribute on the left and a corresponding positive attribute on the right side. The exact questions are presented in Table 4.2. This approach was initially inspired by the QUIS questionnaire by Chin et al. [1988], but was heavily adapted to this research setting. The measurements are presented in Figure 4.6.

Participants rated each countermeasure on six scales

HL was in average rated the worst in each category. *LL* and *HD+M* were rated neutral in all categories, with the exception, that *HD+M* was ranked relatively bad in the *Clear* category. *SW* was also rated positive to neutral, while also being the second most *Helpful* and *Safe* countermeasure. In most categories, *HD* and *HL+E* were rated best, often similarly well. However, in the categories *Safe* and *Helpful HL+E* strongly outperformed *HD*.

In most categories *HL+E* and *HD* were rated best

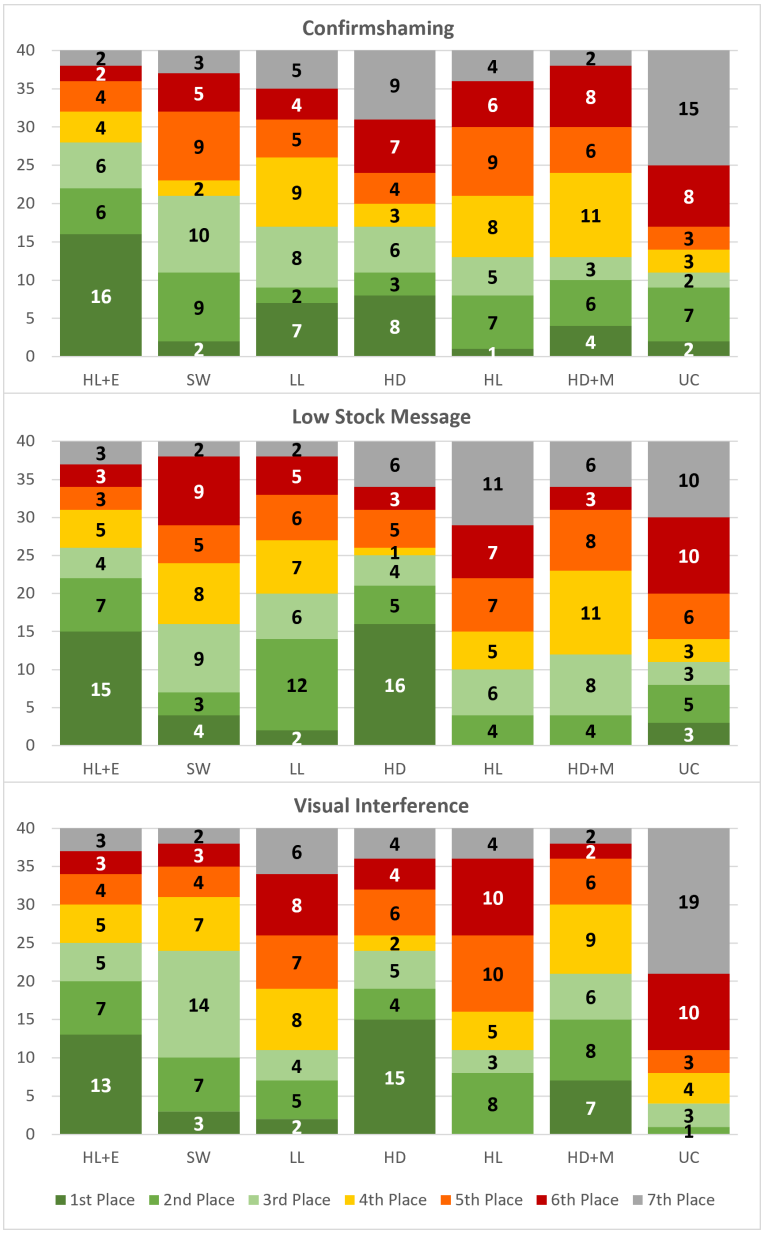


Figure 4.4: The rankings for the individual DPs. For consistency, the order is the same as in Figure 4.3. Throughout all conditions, *Highlight with Explanation* (*HL+E*) received good rankings, while *Unchanged* (*UC*) received bad rankings. The results for the other countermeasures vary.

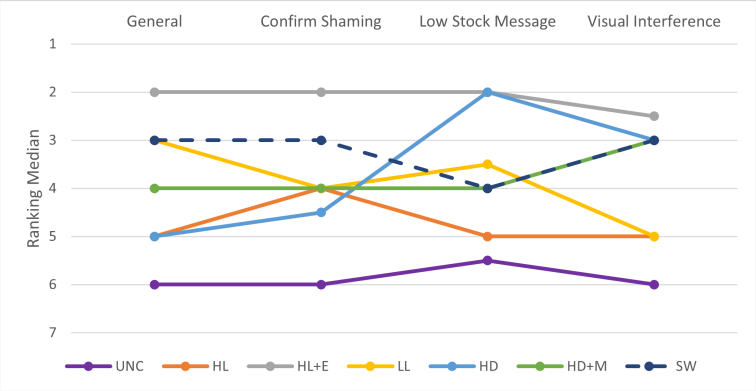


Figure 4.5: This line graph illustrates the differences between the general ranking (on the left) and the DP-related rankings. The order is adjusted for best readability. Each line represents a countermeasure, and each point represents the median of the rankings for that countermeasure. For example, *HD* is ranked much better for *Low Stock Message* than in general.

4.2.4 Participant Comments

In this section, we present the qualitative data that we gathered in Study 1. Numbers in brackets denote how often a remark occurred in similar form. For direct user citations, we provide the randomized participant ID in brackets with a leading *P*: *"This is an example comment"* (*P6*). Participant citations have been translated if necessary and corrected in spelling and grammar.

Many participants encountered DPs on shopping websites

At the beginning of the study, we asked participants how they encountered DPs in their everyday lives. The most common contexts of DP encounters were shopping sites (10) like Amazon (2) or ticket shops (2), cookie banners (9) or shipping methods (7). The DPs most frequently described are instances of *Hidden Costs* (10), *Hard to Cancel* (5), *Misleading Ads* (5) *Visual Interference* (5) and *Forced Accounts/Subscriptions* (4).

We conducted a thematic analysis of participants' comments about the strengths and weaknesses of each countermeasure. This allowed us to explore recurring remarks.

I think using this method, most websites would feel...	
Worse	Better
I think in practice, using this method would be...	
Unhelpful	Helpful
Inefficient	Efficient
Dangerous	Safe
Confusing	Clear
Hard	Easy

Table 4.2: Rating questions answered by participants for each countermeasure on 7-point semantic differential scales.

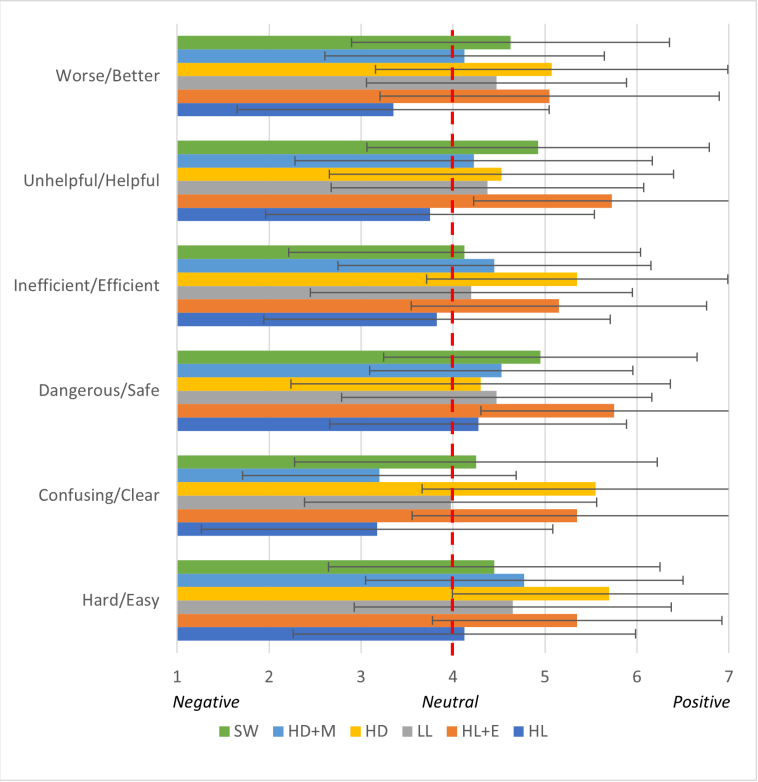


Figure 4.6: Results for the semantic differential ratings on a scale from 1 to 7, with 1 being the negative and 7 being the positive attribute. The red dotted line represents the neutral value of 4. *HL+E* and *HD* received high ratings in most categories. Still, *HD* had one of the lowest ratings regarding safety. The whiskers represent the standard deviation.

Comments on *Highlight without Explanation* (*HL*)

Participants liked that DPs were clearly visible, but wanted additional information

The most frequent critique for *HL* was that there was no additional information or explanation given on why something was highlighted (11): *"Users might need an explanation to decide which is the right option to choose"* (*P6*). Additionally, the highlighting with a red box was perceived as intimidating (2): *"Not even I would dare to click any button"* (*P32*). This is in line with the second frequent remark that *HL* clutters/worsens the website (7): *"I don't like the visual clutter"* (*P11*). Participants were concerned that the red box could strengthen a DP in certain situations (5), defying the purpose of the countermeasure: *"This highlights the text 'Hurry, limited quantity left' so that the user might be even more confused"* (*P3*). Participants liked about *HL* that it alerts users (5) since *"It clearly shows the attempt"* (*P10*), and facilitates user reflection (5) by *"alarm[ing] you to check yourself what the website probably wants you to do"* (*P13*). Participants appreciated the content of the website not being changed (3) so that it is *"transparent"* (*P12*) *"like a helpful overlay"* (*P1*), making it *"very simple"* (*P22*).

Comments on *Highlight with Explanation* (*HL+E*)

Participants liked the explanation, but disliked the visual clutter

Many participants perceived this countermeasure as too visually cluttered (14). The warning sign was perceived as *"patronizing"* (*P6*). *"It looks like a download button for a virus"* (*P32*). Especially on websites with many DPs, *"several such indicators may make the website somewhat confusing and complicated to use"* (*P9*). It was perceived as *"annoying over time"* (*P3*), *"daunting"* (*P17*) or *"confusing"* (*P18*). *P12* suggested adding the option to deactivate certain explanations for the future, to counteract this. Participants liked the fact that there was an explanation given (12), thus educating the user (9) and fostering informed decisions (6): *"You get all the information you need and still get to make the final decision"* (*P1*). This would also help inexperienced users (*P2*). Participants entertained the idea that this countermeasure would *"frustrate"* (*P19*) the manipulating parties, as *"Clear call-outs of scumfuck behaviour"* (*P26*) (3). While two participants liked that the explanations only showed up on hover,

there were concerns if they would be used since *"nobody reads hidden information"* (*P29*). *P6* remarked that *"The learning effect will be enormous, maybe even to the point where this can be turned off again"*.

Comments on *Lowlight* (*LL*)

The participants had mixed feelings regarding the core characteristic of this countermeasure. On the one hand, the effect of manipulation is reduced (8), because *"it robs a lot of common dark patterns of their most useful feature" P40*. On the other hand, manipulation is not stopped completely (8). This applies especially to manipulative text since *"most biased phrasings are still more noticeable"* (*P32*) and *"the evil text might still be visible"* (*P3*). Participants were concerned that *LL* makes the site harder to use (6), especially for people with vision issues (2). Changes made by the countermeasure might be part of the website design (5), which could result in shady websites looking misleadingly trustworthy (2) since *"the manipulation is now even more subtle"* (*P6*). Participants liked that *LL* is visually clear (4) and supports prioritizing information on a site (4) because *"It does not visually interfere that much"* (*P1*). Additionally, it exposes manipulation attempts (3), but does not change the content (3), making it *"much safer and clearer than hiding things"* (*P13*).

Participants were unsure if *LL*'s effect was strong enough

Comments on *Hide without Marking* (*HD*)

Many participants feared that this method could remove relevant content from a page (11). Since it does not indicate changes in any way (10), this *"could lead to really dangerous results [...]"* (*P9*). If the countermeasure worked flawlessly, users could not use the presence of DPs to recognize shady websites (2), since it *"hides the villainous nature of the vendor"* (*P26*). Additionally, this countermeasure *"might lead to censoring"* (*P3*) or *"creators of the countermeasure could use it to manipulate in other ways"* (*P12*). Participants appreciated that this countermeasure *"disarms"* (*P10*) or even *"eliminates the possibility"* (*P9*) for manipulation (11). This results in a clear website without extra content (9), reducing cognitive

Participants like that *HD* disarms DPs, but fear that it might remove relevant content

load for the user (8), and allowing them to *"decide completely unbiased"* (*P2*) (4). Therefore, it *"could be a good solution for parents to protect their children"* (*P36*).

Comments on *Hide with Marking* (*HD+M*)

Highlighting areas where a DP has been removed seems to confuse users

In contrast to *HD*, this method makes clear where something was altered. However, 12 participants stated that it *"might be confusing if hidden information is highlighted"* (*P39*). Additionally, participants were interested in the reason why something was changed (8). Users *"want to know what is hidden and why"* (*P23*), because *HD+M "could make you feel unsafe about a potentially harmless website through the fear of the unknown it creates"* (*P40*). Users suggested that it should be possible to see the original content (6) or that an explanation should be available (3). Note that the former suggestion resembles *SW*. As with *HL*, participants liked that it counteracts the manipulation effectively (7), but additionally shows that manipulation was present (5): *"I am unbiased while still knowing that someone wanted to trick me"* (*P2*). It is a simple/easy countermeasure (3) that lowers the cognitive load (2).

Comments on *Switch* (*SW*)

Participants liked that they had the choice of seeing the original or staying completely unbiased

Participants liked the option to see the original content (11), allowing them to reflect on it (4). They liked that it hides the DP initially (6): *"I am initially unbiased. It isn't distracting. I can see the manipulation as it was intended if I want to"* (*P2*). *"It protects the user from unaware manipulation, as the manipulation becomes only visible after being made aware"* (*P36*). Participants think *"it might be fun to hit the switch"* (*P40*) (2), but there are concerns that it makes surfing inefficient (5) as it clutters the page (3) and users might *"click it all the time to double-check whether the program decided right"* (*P13*). However, the option to switch to the original content allows *"to build up trust with the countermeasure tool: Whenever I still desire to know why the interface was altered, I have access to the original, but over time I will look at that less and less often as I see that the tool only does helpful alterations"* (*P12*).

4.3 Discussion of Study 1

In this section, we discuss the findings of Study 1 and how Study 2 was designed on the basis of these findings.[4]

4.3.1 General Findings

The participants were irritated by DPs and receptive to the idea of DP countermeasures

The participants appreciated the concept of having a browser plug-in that counters DP. *UC* was ranked the lowest in general, as well as for each of the DPs tested. In our qualitative data, we found a strong negative tendency among participants toward DPs and DP designers. Our data also suggest that there are notable differences in the perceptions of the countermeasures we studied.

4.3.2 Information Content of Countermeasures

Countermeasures that provide comprehensible information seem to be preferred by users

Based on the general ranking (Fig. 4.3), it seems like users prefer countermeasures that are information rich. This includes the reasons why something was classified as a DP (*HL+E*), areas on a page that were changed (*HL+E*, *SW*), and the original content (*SW*, *LL*). Nevertheless, any added information should be transparent. For instance, the marking of the area of a DP without including any extra information (*HL*) is unwanted. When a marking only signifies that an alteration has taken place, but not how (*HD+M*), users perceive it particularly negatively. Although *HL+E* was often ranked in first place, 14 out of our 40 participants stated that it creates too much visual clutter. It is surprising that participants perceived *HL+E* and *HD* as making pages feel better to a similar degree, since *HL+E* introduced the most visual clutter, and *HD* not only introduces no additional elements, it even removes elements from the existing page. Therefore, one of our goals in Study 2 is to investigate how users perceive the visual clutter caused by additional information during real interaction.

[4]The findings of Study 1 will be published in our recently accepted paper "Investigating Visual Countermeasures Against Dark Patterns in User Interfaces" by Schäfer et al. at MuC'23.

4.3.3 Hiding DPs with Countermeasures

For *HD*, some feared that relevant content would be changed unnoticed, others enjoyed that the DPs were gone

At first glance, *HD* might seem like an obvious solution against the manipulative effects of DPs, and we expected it to be popular among participants. Surprisingly, this was not the case. It was ranked very controversially by participants in the general ranking (Fig. 4.3) and was rated as the second least safe. Participants were concerned that relevant content could be inadvertently corrupted without being apparent. Several participants expressed concern that *HD* would make sketchy websites appear deceptively trustworthy. Initially, we expected that marking affected areas with *HD+M* would be a less dubious version of *HD*, but *HD+M* was rated far worse than *HD* in every category (Fig. 4.6), and barely ranked better than the baseline *UC*. Participants who trust technology might be more comfortable with *HD*. It was rated as the simplest, clearest, and most effective countermeasure, and participants liked that it would completely remove the manipulative effect of DPs if it worked reliably.

SW is a less dubious but also less efficient version of *HD*

SW seems to address the problems of *HD* by allowing you to see the original view. It was ranked better in general and was rated the second safest, but it was also rated significantly worse in all of *HD*'s strong categories. It may distract users from their goal on the site, because they want to see what elements have been hidden from them.

Not all DPs can be removed

It is important to note that some DPs cannot be removed under our basic assumptions made in Section 3.4. *Forced Enrollment* is one example of that. We investigate this closer in Study 2, where we tested almost all DP types from the selected taxonomy (Tab. 2.1).

4.3.4 Mixed Approach Countermeasures

Different approaches can be combined based on DP type and user preferences

Two factors seem to influence the selection of a suitable countermeasure: the type of DP, and the user's preferences. It seems useful to combine different approaches. Patterns that grab the user's attention without providing relevant information, such as *Low Stock Messages*, could be countered with *HD*, especially if *HL+E* would result in too much visual clutter. *HD* received its best ranking for this DP. If

content could be relevant, *HL+E* can be used, it was ranked best in general and good in all individual rankings. The countermeasure preferences for different DPs are examined in more detail in Chapter 5. User preference is likely to be influenced by DP awareness and confidence in the system. Both of these factors can be expected to change over a period of time while the countermeasure is in use, so it should be adjustable or self-adjusting over time.

4.4 Limitations of Study 1

In the following, we briefly address the limitations and compromises of Study 1. Since most of our participants have a technical background, they may not be representative of the average user. We were able to test six countermeasures for three DPs. Other probable countermeasures and opposing mechanisms for contracting DPs, such as laws, were not evaluated in our study. Finally, our study did not include interaction with a website, but was entirely based on static screenshots. Therefore, we opted for a lab study using interactive prototypes in Study 2.

The goal of Study 2 is to address the compromises from Study 1

Chapter 5

Study 2: Lab Study

Study 2 explored how users perceive countermeasures in a realistic setting that involved multiple DPs on a single page and a real interaction process. This chapter begins with our research question and hypothesis. Then, we provide a brief description of the creation of our prototypes and the countermeasures that we investigate in this study. Afterward, we explain our study design and procedure. Finally, we present our quantitative results in detail and summarize our qualitative results. Our findings are discussed in Chapter 6.

5.1 Research Question and Hypothesis

Study 2 compares fewer countermeasures in a more realistic setting

Study 2 augmented Study 1 in the following ways:

- Study 2 was a lab study, not an online study.
- We used interactive prototypes instead of screenshot prototypes.
- The setting was more realistic, with multiple DPs per page instead of only one.
- Users could make decisions during the interaction process.

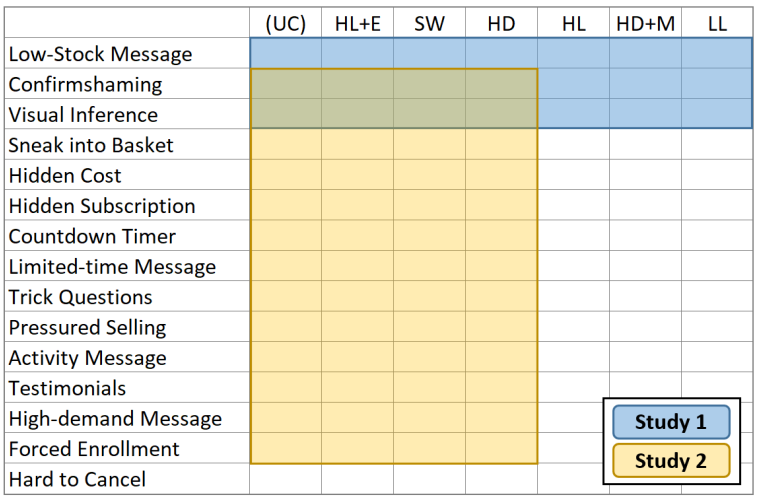


Figure 5.1: Combinations of DP types and countermeasures that we tested in Study 1 and Study 2. The DP types are taken from our selected taxonomy (Tab. 2.1). Study 1 focused more on comparing many countermeasures and investigating whether user preferences changes for different DPs at all. Study 2 focused more on testing some CM for as many DP types as possible.

- We increased the number of DPs we investigated, but reduced the number of countermeasures (Fig. 5.1) based on findings from Study 1.

Our research question for this study was:

RQ: *How do different countermeasures impact the interaction with pages that use multiple DPs?*

We hypothesize

- **H1.1**: Different countermeasures will have different usability.
- **H1.2**: Some countermeasures will enhance the usability of a page.
- **H2.1**: Different countermeasures will affect the efficiency of the user.

- **H2.2**: Some countermeasures will enhance the efficiency of the user.
- **H3.1**: Certain countermeasures will be preferred in general.
- **H3.2**: Certain countermeasures will be preferred for certain DPs.
- **H4**: Countermeasures will impact how trustworthy a page feels.

5.2 Prototype

We created two interactive shopping scenarios

We designed interactive prototypes using the prototyping tool Figma[1]. In Study 1, online shops and especially ticket shops were mentioned by participants as typical examples for websites that involve many DPs. So we created two different website prototypes: a *Smartphone Shop* and a *Ticket Shop*. They are both common shopping scenarios, while having different typical visual appearances. Each page consisted of two to three views and included 7 DPs from the taxonomy of Mathur et al. [2019] (Tab. 2.1). The implemented DP instances were adapted from examples in their work where possible. As an additional example prototype, we created a simple *Muffin Store* with only one view. This was used as a small practice example at the beginning of each condition. The prototypes are shown in Appendix C.

5.3 Countermeasures

Study 2 examines *HL+E*, *HD*, *SW*, and *UC*

We chose to examine a subset of the countermeasures from Study 1 because this allowed us to compare the corresponding results. We decided on four countermeasures, including a baseline. This resulted in a reasonable study duration of one hour and a manageable number of combinations. Based on the online study, we included the countermeasures *Highlight with Explanation* (*HL+E*) and *Switch*

[1] `https://www.figma.com` (accessed march 2023)

(*SW*) because they received the best overall rankings. Since *Hide without Marking* (*HD*) received the most controversial overall ranking and was ranked well for some specific situations, we included it. The countermeasures were implemented as described in Section 3.4. The only adjustment we made based on participant feedback from Study 1 was to change the icon of *SW* from a light bulb to a toggle button to make it clearer.

5.4 Study Design

We conducted a within-subjects-like study with a custom counterbalancing strategy

The design of Study 2 is derived from a within-subjects design with four conditions. We extended this design approach to meet our requirements: To counterbalance the conditions, a four times four Latin square was used. We expected a large learning effect across the study, so we created two different prototypes (*Smartphone Shop*, *Ticket Shop*) that used different sets of DPs. An alternating order was chosen for the websites to minimize the learning effect. When conducting the study multiple times, we would first perform one Latin square iteration, starting with the *Smartphone Shop*. Afterward, we would repeat it starting with the *Ticket Shop*, resulting in a total of 8 participants per iteration. We consider this to be the best balance between counterbalancing, minimizing learning effect and resource effectiveness. Additionally, some measurements, such as task completion time, were only measured for the first appearance of *Smartphone Shop* and *Ticket Shop*.

5.5 Study Procedure

Study 2 consisted of two parts. The first part was similar to Study 1 and focused on the characteristics of the different countermeasures, while the second part aimed to compare the countermeasures for different DPs. A structural overview of Study 2 is presented in Figure 5.2. The questionnaire is included in Appendix B.

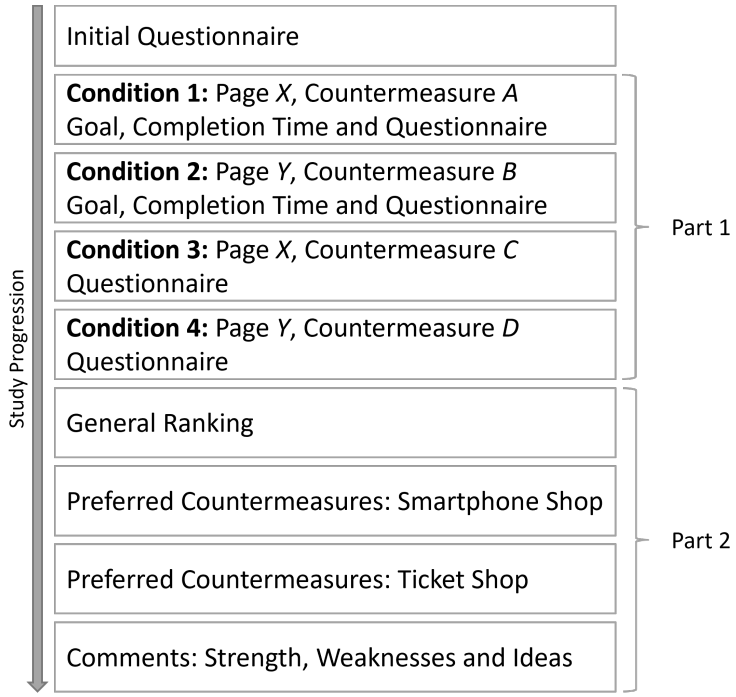


Figure 5.2: The structure of Study 2. Participants performed one task and completed one questionnaire in each condition. Condition 1 and 2 involved more measurements since they were the first time the participant encountered the website. *X* and *Y* are different elements from {*Smartphone Shop*, *Ticket Shop*}. A to D are different elements from {*UC*, *HD*, *SW*, *HL+E*} in orders taken from a Latin square.

5.5.1 Part 1: Individual Countermeasure Characteristics

Participants performed tasks on websites with different countermeasures

In the first part, the current countermeasure for each condition was explained to the participants. After that, the example *Muffin Shop* with the current countermeasure was presented to the participants. In addition, a print-out from the unchanged *Muffin Shop* was given to the participants for comparison. Participants were encouraged to take their time and ask questions during this step. When they were finished and felt that they understood everything, the actual task with the current countermeasure applied was pre-

sented to them. They were asked to perform the task in a way that felt efficient and natural to them. The different tasks were:

- **Muffin Shop** *(example)*: Order muffins for an office party. This is an example, so you can see how the method works. Take the time you need.
- **Smartphone Shop**: Order a new smartphone for your 12-year-old cousin. Please solve this in an efficient manner that feels natural to you.
- **Ticket Shop**: A friend told you about a concert you didn't know about. You want to check it out. Order a ticket. Please solve this in an efficient manner that feels natural to you.

Participants defined their goal themselves, the task provided context to guide them

After completing the task, participants were asked to describe their goal and decisions and were given a questionnaire that included:

- A question about how easy it was for participants to achieve their goal and a question about how confident they were that they had actually achieved their goal.
- A question regarding how trustworthy the page felt.
- The System Usability Scale (SUS) questionnaire (Brooke [1996]) adapted to fit the websites.
- Questions regarding the countermeasure itself (adapted from Study 1).
- A box for further comments.

This was repeated for each of the four conditions.

5.5.2 Part 2: Countermeasure Ranking & Preferences

After participant completed all four conditions, they were asked to rank the methods and justify their ranking. Next,

they were asked to choose which countermeasure they would prefer for each DP on the two websites, which countermeasure they would prefer. Finally, they were asked for their opinion regarding the strength and weaknesses of the countermeasures and for suggestions for improvement.

We collected the preferred countermeasure for each DP, with justification

5.6 Results

In this section, we present the results of Study 2. We will first display the quantitative results starting with characterizing our participants, followed by the ranking, usability scores, ratings, task completion times, success rates and success easiness/certainty. After that, we present results on preferred countermeasures for specific DPs and qualitative comments.

5.6.1 Participants

We conducted the study with 20 participants (11 male, 9 female) with a mean age of 25.8 years (min = 21, max = 31, std = 2.48). 18 of them completed the study in German, two in English (the spoken language and the questionnaires were adapted, the websites were in English). All participants had an academic background, 19 of them in STEM fields. Four of the participants had previously participated in Study 1. Participants' self-reported awareness of DPs was relatively high (mean = 5.45, std = 1.63, on a scale from 1 "Not aware at all" to 7 "Very aware").

20 participants, all with technical education

The participants were not paid a compensation for taking the study, but each participant could choose a bar of chocolate or vegan chocolate to take home. Additionally, a €20 voucher for a local bookshop was raffled.

Participants were not paid, but received chocolate

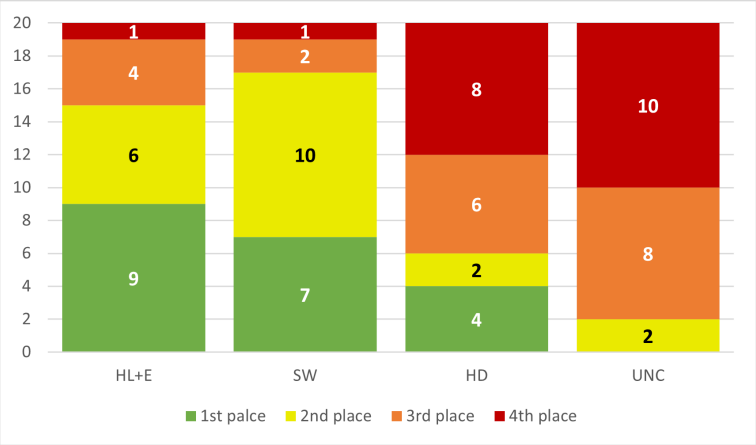


Figure 5.3: Ranking of the countermeasures in Study 2, sorted by average rank. *HL+E* and *SW* have the same average rank and are ranked the best. *UC* is ranked the worst. (n = 20)

5.6.2 Ranking

Overall *HL+E* was ranked best, but the rankings varied widely

Similar to Study 1, we asked participants to rank the countermeasures near the end of the study (Fig. 5.3). In this ranking *HL+E* was placed first the most, followed by *SW* and *HD*. *UC* was not ranked first by any participant, but it was ranked last the most. For each method, every other method has been ranked higher by some participants (Tab. 5.1). 8 participants preferred the baseline (*UC*) over *HD* for the other countermeasures, only 2 participants did. *HL+E* and *SW* were preferred over *HD* by 15 participants, respectively. *SW* and *HL+E* were preferred over each other by exactly half of the participants.

5.6.3 Usability of the Countermeasures

The SUS questionnaire was adapted to the study setting

During the first part of our study, participants received an adapted SUS questionnaire. In our questionnaire, the word system was replaced with the word website. The questions were translated for the German version of the study. The average SUS values are presented in Table 5.2. *HD* and *SW*

Ranked below	*UC*	*HL+E*	*HD*	*SW*
UC	0	18	12	18
HL+E	2	0	5	10
HD	8	15	0	15
SW	2	10	5	0

Table 5.1: Overview over which countermeasures where ranked higher than others. For example 8 participants preferred *UC* over *HD*.

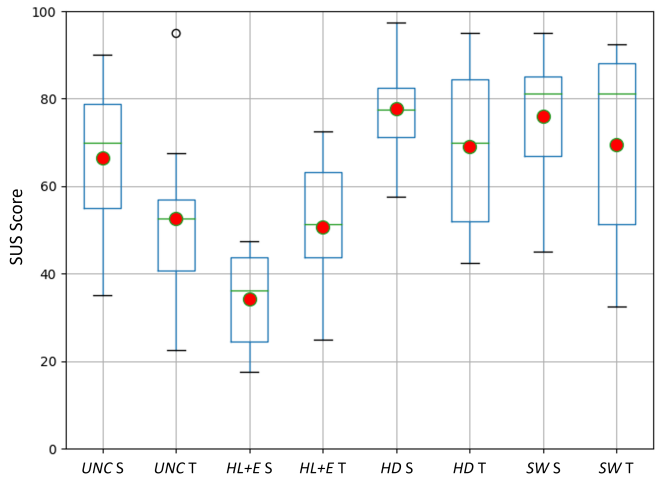


Figure 5.4: Boxplot of the SUS scores for the different countermeasures and websites. The average is represented by a red dot. (S) stands for the *Smartphone Shop*, (T) stands for the *Ticket Shop*. *HL+E* is the only countermeasure that performed worse on the *Smartphone Shop*. (n = 10)

performed quite similar. *HL+E* and *UC* performed similar on the *Ticket Shop*, but there was a huge difference for them on the *Smartphone Shop* (Fig. 5.4).

The average SUS score varies between countermeasures

In direct comparison to the baseline (*UC*), *HL+E* decreased the SUS score by an average of 10.5 (n = 10, std = 26.1) and *SW* increased the SUS score by an average of 15.3 (n = 10, std = 12.1). Due to the counterbalancing design, the data does not allow for a direct comparison between *UC*

	Smartp.	*Ticket*	Avg.
UC	66.5	52.5	59.5
HL+E	**34.3**	50.8	42.5
HD	77.8	69.0	73.4
SW	76.0	69.5	72.8
Avg.	63.6	60.4	62.0

Table 5.2: The average SUS scores by method and page. The SUS score for *HL+E* on the *Smartphone Shop* is surprisingly low. (n = 10)

and *HD*, but it is possible to compare the SUS scores for each page for the two methods. This results in average SUS scores that are 13.9 higher for *HD* than for *UC* (n = 20, std = 22.4).

5.6.4 Perceived Website Trustworthiness

Countermeasures seem to influence the trust in a page

In Study 1, participants often remarked that different countermeasures could change how trustworthy pages are perceived by users. For this reason, we asked participants in Study 2 to rate how trustworthy the pages felt to them. They answered using a semantic differentials scale that ranged from 1 (Very dubious) to 7 (Very trustworthy). In average, *HL+E* was the only countermeasure that made websites seem less trustworthy than the baseline (Tab. 5.3). *HD* and *SW* increased the perceived trustworthiness (Fig. 5.5).

	Smartp.	*Ticket*	Avg.
UC	3	3.33	3.17
HL+E	2.7	2.9	2.8
HD	4.4	4.8	4.6
SW	4.44	4.2	4.32
Avg.	3.64	3.81	3.72

Table 5.3: The average rating for trustworthyness of the pages. The scale ranges from 1 (Very dubious) to 7 (Very trustworthy). (n = 10)

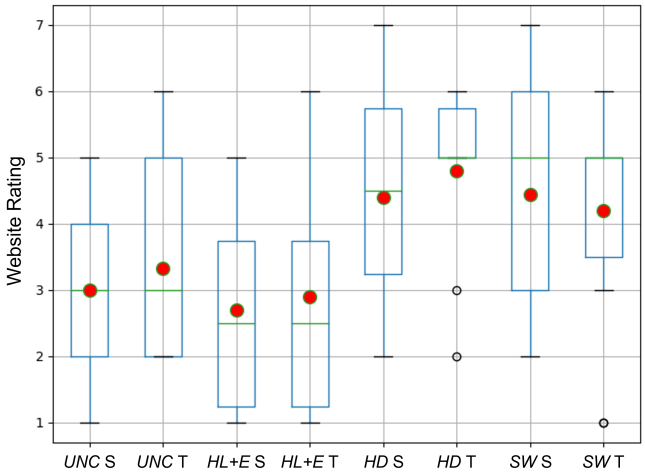


Figure 5.5: Boxplot of the rating of the trustworthyness of a website. The scale ranges from 1 (Very dubious) to 7 (Very trustworthy). The average is represented by a red dot. (S) stands for the *Smartphone Shop*, (T) stands for the *Ticket Shop*. (n = 10)

5.6.5 Task Success

To gain insight into the effectiveness of different countermeasures, we measured task success, task completion time, how easy it felt to reach a goal, and how confident participants were that they had reached their goal. These measurements were taken only for a participant's first encounter with a new website, that is, in the first two conditions (Fig. 5.2). We asked participants to describe their specific goals and decisions after completing a task. This reduced the risk of biasing their intentions. If participants did not fully achieve their goals, for example because they made mistakes or incorrect assumptions, this was counted as an unsuccessful attempt. During the study, we counted 32 successful attempts and 8 unsuccessful attempts, resulting in an overall success rate of 80%. The success rate seems to be influenced by the combination of countermeasure and website (Tab. 5.4).

Participants defined their own goals

HD and *SW* had a lower success rate on the *Smartphone Shop* than *UC*

	Smartp.	*Ticket*	Avg.
UC	100%	**40%**	70%
HL+E	100%	**60%**	80%
HD	**80%**	100%	90%
SW	**60%**	100%	80%
Avg.	85%	75%	80%

Table 5.4: The success rates for the different combinations of website and countermeasure vary widely. The *Ticket Shop* baseline had the worst success rate. *HD* and *SW* seem to decrease the success rate in the *Smartphone Shop*, which is surprising. (n = 5)

All errors made by participants were triggered by DPs

The errors that occurred were:

- *HD*, *Smartphone Shop*: Misconception that a warranty which was automatically added to the basked has to be bought. (1x)
- *SW*, *Smartphone Shop*: The warranty was not perceived and bought by accident. (2x)
- *UC*, *Ticket Shop*: Certain visually inconspicuous buttons were not perceived, resulting in buying an undesired option. (1x)
- *UC*, *Ticket Shop*: Misconception that certain visually inconspicuous buttons were inactive, leading to the wrong conclusion that the desired option would be unavailable. (1x)
- *HL+E*, *Ticket Shop*: A trick question misled the participant to subscribe to a newsletter. (2x)
- *UC*, *Ticket Shop*: Same as above. (1x)

5.6.6 Success Confidence and Easiness

We asked participants how confident they were that they had achieved their goals and how easy it felt (Fig. 5.6). Participants reported in average higher scores of easiness

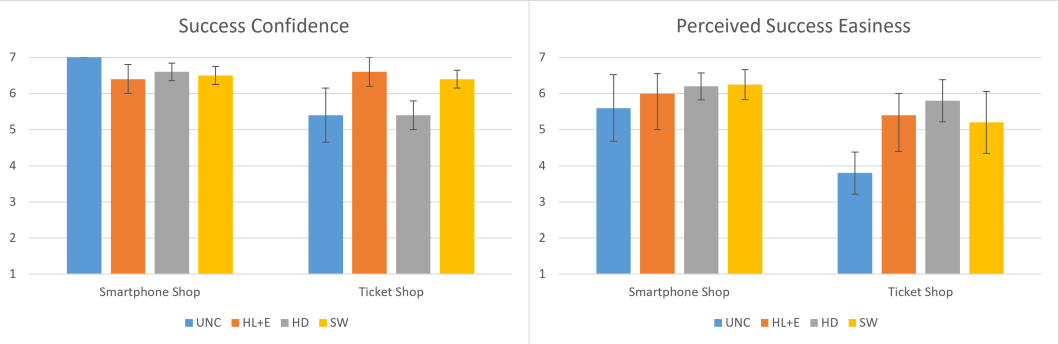


Figure 5.6: The rating of participants about how easy it was for them to reach thier goal, and how confident they were that they did. The scale ranges from 1 to 7, with 7 being the most positive awnser. The black whiskers represent the standard deviation. (n = 5)

Participants overall reported high confidence in success

on the *Smartphone Shop*. On the *Ticket Shop* they reported higher easiness scores for the countermeasured conditions than for the baseline (*UC*). Note that confidence in success on the *Ticket Shop* was comparatively low for *HD*, even though it had a success rate of 100%, while confidence was comparatively high for *HL+E*, which had a success rate of only 60%.

5.6.7 Task Completion Time

HD was slower than *UC*

The task completion times are plotted in Figure 5.7. Surprisingly, *UC* had a smaller completion time than *HD*. *HL+E* seems to drastically increase the completion time for the *Ticket Shop*, while it did not change much for the *Smartphone Shop*.

5.6.8 Rating of Countermeasures

Countermeasures in each category were rated positively on average

Participants rated the countermeasures on the same seven point semantic differential rating scales as in Study 1 (Tab. 4.2). The results are presented in Figure 5.8. All average ratings are in the positive range above 4. *HL+E* was rated on average as the safest, but also the least likely to increase perceived page quality. *HD* was on average rated worst

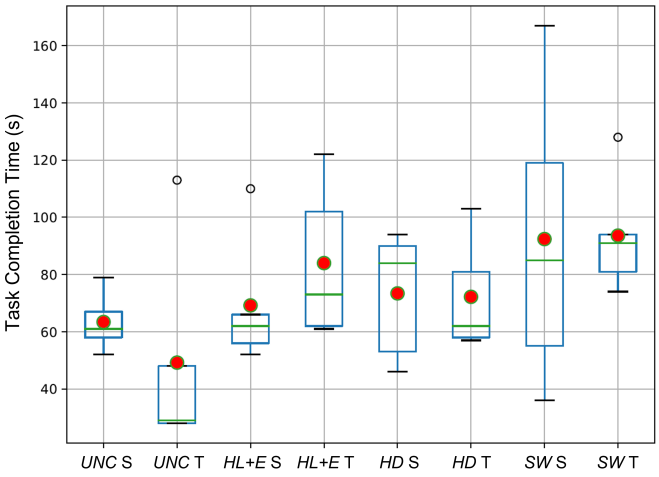


Figure 5.7: Boxplot of the task completion times in seconds. The average is represented by a red dot. (S) stands for the *Smartphone Shop*, (T) stands for the *Ticket Shop*. (n = 5)

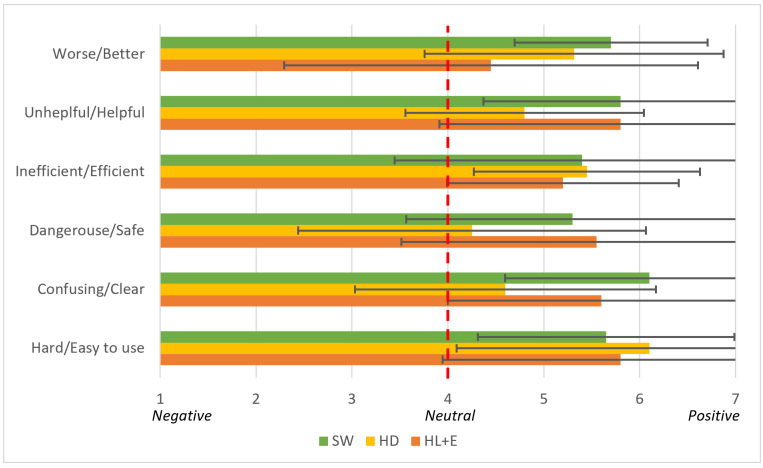


Figure 5.8: Average ratings of participants for different aspects of the countermeasures. The scale ranges from 1 to 7, with 7 being the most positive. The red dashed line marks the neutral value 4. The standard deviation is represented by the black whiskers. (n = 20)

for being clear, helpful and safe, but best for being easy to use. *SW* was rated on average as being the clearest and increasing page quality the most. It was not rated worst in any category.

5.6.9 Preferred Countermeasures for Different Elements

We asked participants which countermeasure they preferred for each DP and why. An overview of the different DPs used in this study is given in Figure 5.9. The results of what the participants chose are presented in Figure 5.10. Below, we present the most common reasons why each countermeasure was preferred for a given DP. Next, we summarize the results for the different DPs. To provide a more comprehensive overview, we have grouped the DPs based on their goals and the countermeasures selected for them.

We asked participants to choose their favorite countermeasure for each DP

General Reasons for Countermeasures

The total number of justifications provided for countermeasure preferences is indicated by numbers in parentheses. Theoretically, the upper limit for this number is 520 (40 participants times 13 DPs).

The most frequently selected countermeasure was *HL+E*, it was chosen in 40% of all cases. The most common reasons were that:

Participants preferred *HL+E* in 40%, *HD* in 35%, *SW* in 20% and *UC* in 5% of the cases

- It provides additional information (38).
- It warns the user (24).
- It highlights relevant aspects of the website (22).
- It does not change content (that might be relevant) (17).

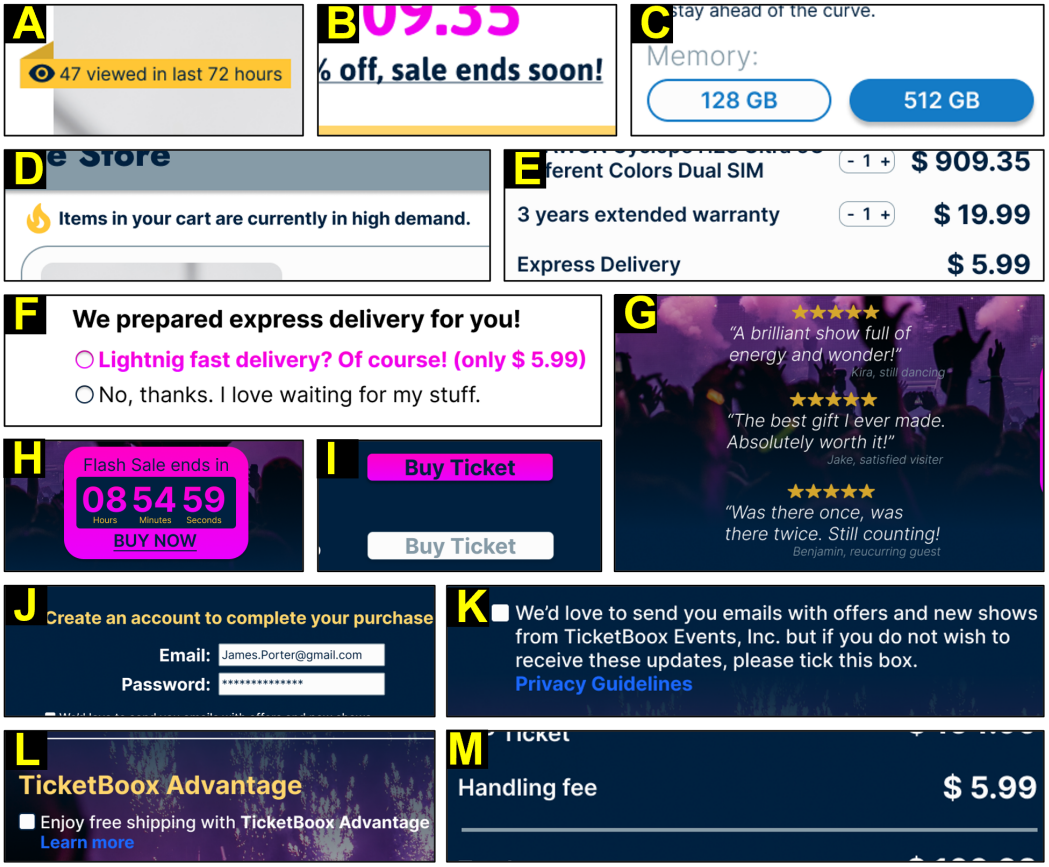


Figure 5.9: The DPs tested in the study. A to F are part of the *Smartphone Shop*. G to M are part of the *Ticket Shop*. Based on the taxonomy (Mathur et al. [2019]) they can be classified as follows: A) *Activity Message*, B) *Limited Time Message*, C) *Pressured Selling*, D) *High Demand Message*, E) *Sneak Into Basket*, F) *Confirmshaming*, G) *Testimonials*, H) *Countdown*, I) *Visual Interference*, J) *Forced Enrollment*, K) *Trick Question*, L) *Hidden Subscription*, M) *Hidden Costs*

HD was chosen in 35% of all cases. The most common justifications were that:

- The element does not contain relevant information (26).
- The element itself is not relevant (21).
- It cleans the user interface (18).
- It gives users more control for an unbiased decision (14).

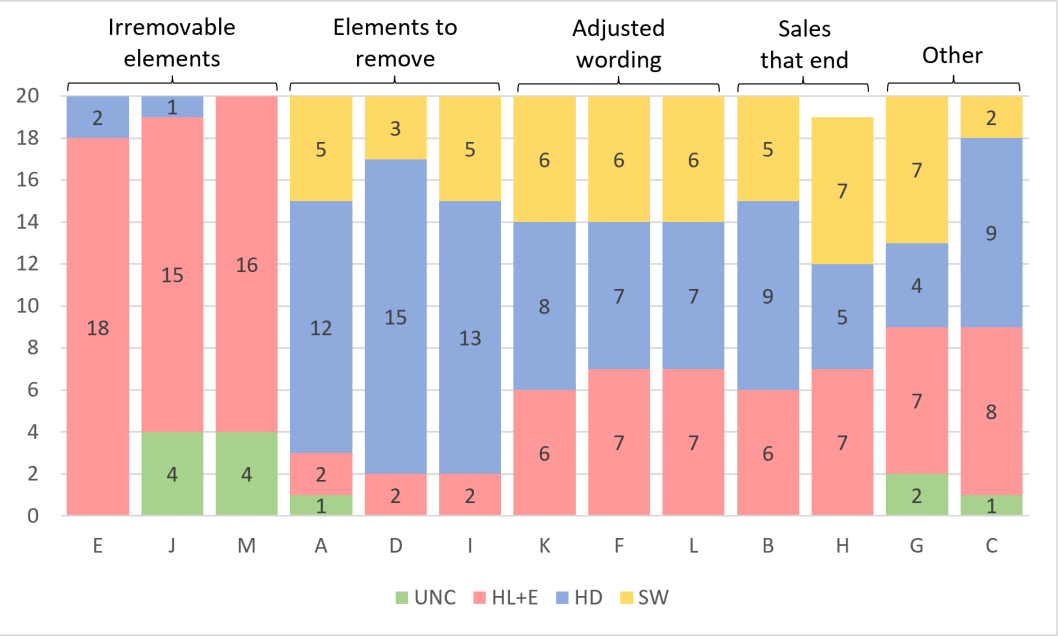


Figure 5.10: The preferred countermeasures for the various manipulative elements in our study. The elements are labeled according to the order in which they appear on the page. In this diagram, the elements are grouped according to their characteristics. Screenshots of the elements and which DP they correspond to are shown in Figure 5.9.

Participants chose *SW* in 20% of the cases, primarily because:

- It allows users to see the original element (23).
- It indicates that something was changed in a certain area (12).

In the following, we present the most relevant justifications for the preferred countermeasures against the 13 individual DPs of our study. The notation (A:2, B:6) is used to indicate that a justification was made 2 times for element A and 6 times for element B.

Irremovable Elements (E, J, M)

The use of *HL+E* was preferred for elements that cannot be removed

In our study, we included three manipulative elements that could not be counteracted by *HD* or *SW*. These elements are instances of *Sneak Into Basket* (*E*), *Forced Enrollment* (*J*) and *Hidden Costs* (*M*). For those elements, *HL+E* was clearly preferred. Participants said that they liked the additional information (E:7, J:8, M:3), that it highlights relevant aspects of a page (E:7, J:2, M:4) and that it warns the user (E:8, J:3, M:5).

Elements to Remove (A, D, I)

The use of *HD* was preferred for content that is irrelevant or misleading

For representations of *Activity Message* (*A*), *High Demand Message* (*D*) and *Visual Interference* (*I*), participants clearly preferred *HD*. While the reasons for this were quite similar for *A* and *D*, they were rather different for *I*. The biggest arguments for *HD* on *A* and *D* were that the element has no useful information (A:7, D:5, I:1) and that it is not relevant (A:6, D:5, I:1). The biggest argument for removing *I* was, that it was misleading (A:0, D:0, I:5) and that removing it cleans the UI (A:1, D:2, I:3). This is well in line with the fact that A and D work very similar, but very different to *I* (Fig. 5.9).

Adjusted Wording (K, F, L)

There was no clear preference for DPs that primarily use wording tricks

For the following elements, there was no clear priority for a single countermeasure. The elements representing *Trick Question* (*K*), *Hidden Subscription* (*L*) and *Confirmshaming* (*F*) were all balanced between *HD*, *SW* and *HL+E*. Additionally, *F* applies *Visual Interference* on the text. *HD* was preferred because it cleans the UI (F:3, K:4, L:3). Especially for *F*, participants felt like it would give users the control for an unbiased decision (F:4, K:0, L:2). *SW* was preferred, since it has the option to see the original (F:4, K:2, L:1). This seems to be especially important for *F*. The changes by *SW* also seem to make the page more pleasing/better to use (F:1, K:2, L:2). For *HL+E*, participants liked that it highlighted

relevant aspects of a page (F:1, K:3, L:1), especially for *K*, which was an easy to miss *Trick Question*.

Sales that End (B, H)

The DPs representing *Limited Time Message* (*B*) and *Countdown* (*H*) work very similar. Both aim at creating a sense of urgency during shopping. However, *HD* was the most preferred method for *B* and *HL+E* and *SW* were preferred more for *H*. It seems like participants chose *HD*, since the information was not useful (B:4, H.3), but for *B* additionally because it was irrelevant (B:3, H:1), nonsense/annoying (B:2, H:0) or not interactive (B:2, H:0). On the other side, *SW* was preferred for *H* by some since its information might be relevant (B:1, H:4) and because it neutralizes the manipulative effect (B:0, H:2). Others decided for *HL+E* as the content might be important (B:3, H:5) or because *HL+E* provides additional explanations (B:1, H:3). For both elements, participants chose *SW* if they wanted the option to see the original content (B:3, H:2).

For *B*, *HD* was chosen more than for *H*, since the information is less specific

Reviews (G)

The element *G* represented the *Testimonials* DP. *HL+E* and *SW* were both most popular, followed by *HD* and also *UC*. *SW* was chosen by participants that wanted to use the option to see the original (5) and an indicator that something was changed in this area (2). Others preferred *HL+E* since it does not change the content of the element (4) and provides additional explanations (2). While some participants who chose *HD* seem to think that the testimonials were irrelevant (3) or did not provide useful information (2), two participants also chose *UC*, saying that they would not have noticed them anyway (2), that they were useful (1) or that a removal looked bad (1). It seems like in this case, participants had different opinions whether this element is a manipulative fake or relevant information. In the study, they were only told that the *Testimonials* have a questionable origin.

Most participants wanted to be able to see the *Testimonials* in some way

Choose the Expensive Option (C, I)

Each of the websites in the study included an element that tried to nudge users towards more expensive options. One was *Pressured Selling* (*C*), which means that a more expensive version was pre-selected. The other one was *Visual Interference* (*I*). This means that some options were presented visually different, in this case to make the more expensive option more attractive. It is interesting to compare these two elements, because although they look very similar to users, the implications of removing them differ. It does not need much intervention to remove visual interference, but a pre-selection can not be removed completely. The only thing that can be done on client side is to remove the visual representation of the pre-selection, not the pre-selection itself. This results in a mismatch of the visible state of the page and the true state of the page. Some participants disliked this. *I* has been one of the elements where *HD* was clearly preferred (as described in Section 5.6.9), *HD* and *SW* were both almost equally popular for *C*. However, the reasons for choosing *HD* were different for the two elements. For *C*, participants felt like it would give them more control (C:5, I:3), while for *I* the element was deemed misleading (C:0, I:5) and the website was viewed to be cleaner (C:0, I:3). *HL+E* was also popular for *C* because it provided additional information (C:5, I:2) and highlighted an important aspect of the page (C:4, I:0).

It is technically hard to remove pre-selections on the client side

5.6.10 Qualitative Comments on Countermeasures

All comments about countermeasures examined in Study 2 or countermeasures in general were collected. Using thematic analysis, they were classified as strengths, weaknesses, suggestions, or general comments. In this section, we first provide a summary of the comments for each countermeasure, followed by a summary of general remarks. Numbers in brackets indicate how often a particular type of comment was made. When quoting users, we include a randomized participant ID with a leading P and the number in brackets.

Comments on *Unchanged*

Several participants were bothered by *UC*, but a few appreciated that the page remained unmodified

In the *UC* countermeasure, which was the baseline of our study, 13 participants reported the weakness that here the manipulation is completely present / users might get biased. Some see this as a danger for themselves, others mainly for others. *"[UC] only makes sense with a good knowledge of how to use the internet"* (*P4*). *"One thinks one is experienced, but the manipulation still works too often"* (*P10*). The unchanged pages felt more annoying/hostile (4) but also safe, since nothing is changed (4). Participants liked that the information is preserved (4).

Comments on *Highlight with Explanation*

The combination of highlighting and additional information provided by *HL+E* was well received

For *HL+E*, the participants liked that it provides useful information and explanations (11). The boxes also warn the user and make the DPs easy to spot (10). 7 participants also liked that this method could increase the awareness of users. This could make them less prone to manipulation by DPs in general (2). *"The explanations are nice, I enjoy learning something. If I used this method frequently, I would likely start ignoring the texts"* (*P8*). Other positive aspects were, that this method highlights all the DPs (3), that it exposes the manipulation attempts clearly (4), that it does not change content (2) and that it feels generally safe to use (3). *"[The] website feels worse, but that is ok since if this variant [countermeasure] has this effect it [the website] has not deserved otherwise"* (*P10*). *"Because of the many warnings, the page seems less trustworthy and makes me question the page more"* (*P16*). These comments also already hint at the largest point of criticism on this method. 10 participants remarked that the website gets visually cluttered/overwhelming or confusing. Even worse, 3 participants remarked, that this method might focus too much attention on DPs. This could strengthen their effect, resulting in exactly the opposite outcome we desired. *"The red dotted box and the exclamation mark seem dangerous, might focus too much attention on those elements"* (*P8*). Participants were also not sure whether a warning alone would stop the manipulation (4). And some said that it is *"cumbersome and time-consuming"* (*P18*) to read the explanations

HL+E might draw too much attention to some DPs, making them more effective

(4). 3 participants said they disliked the visual appearance. 2 participants commented that it looks like part of the website itself. *"[...] I felt like this would be another element that is part of the website that tries to trick me somehow [...]"* (*P9*). *P18* suggested making the box less dramatic, and *P5* suggested making them customizable. It seems also not desired by everyone that the boxes are always present. *P2* and *P14* wanted a switch to turn them off, and *P15* suggested to only use it in hard cases.

Comments on *Hide without Marking*

HD creates a feeling of intransparency

HD adds no visual elements to the view at all. Many participants stated that it was (very) clean, quick or simple (9) or that the page was easy/nice to use (5). *"I did not feel the existence of the implemented method at all"* (*P11*). *"The website felt very easy and clear"* (*P20*). Participants also acknowledged that this method would (completely) hide/reduce the manipulation, resulting in a strong protection (8). While one participant said that this method would feel safe, many participants said that the method would feel (very) unsafe or dangerous (8). The user cannot know/control what has been changed (7), but the method could change/remove relevant information (6). This might also result in the risk to make sketchy pages look more trustworthy than they actually are (5). *"Simply removing the DPs could make some websites look too trustworthy so that one could easily fall for a scam"* (*P8*). Some of the negative aspects pointed out were also related to the users. Some participants disliked that the method could not counteract all DPs (3). This is a problem because users might overtrust the method (2). *"Removing misleading things is nice, but it poses the risk to produce a false sense of safety and one might miss things that could not be removed"* (*P7*). It would also not increase the awareness of users on DPs (2).

Comments on *Switch*

What was liked most on the *SW* countermeasure was the option to compare the changed element with the original

element (10). Participants saw different advantages in this. For example, users can adjust what is shown to them or decide what to inspect closer (4). *"[SW] gives me confidence, since I have the switches as indicators and by toggling I can see how trustworthy the page is"* (*P12*). It also allows users to check if the countermeasure works correctly (3), learn something about DPs (3), and estimate how trustworthy a page actually is (2). The countermeasure also makes the page easier to use (5) and the user can see where something was changed (5). It reduces or hides manipulation (4), while keeping the original content accessible when needed (4), resulting in an unobtrusive or visually pleasing method (4). *"[SW] gives me the possibility to quickly make purchases while still being informed about removed misleading content and learning from it which patterns exist to create an awareness on them and being able to recognize them by myself"* (*P12*). A drawback that was seen the most with this method was that it might make users less efficient (6). It is cumbersome to use because of the many clicks (4), it does not provide explanations (4), and makes users curious on what is behind the switch (3). *"The switches are kinda fun to play with, but also a bit distracting and overcomplicate the website"* (*P11*). Participants also felt like the additional switches would clutter the page too much (5). Based on this, 6 participants suggested to include a global switch for all DPs (maybe leaving out the single instance switches completely). Other suggestions included a short tutorial-like explanation on how the switches work (1), to activate the switches on hover instead of on click (1), or to use the hovering similar to *HL+E* to include additional tooltips (1).

SW gives users more control, but it may make them less effective

Many participant asked for a global switch to toggle all DPs at once

General Suggestions and Remarks

We asked people for ideas to improve the countermeasures presented, and by far the most common response was to combine *HL+E* and *SW*. There seem to be many different ways to do this. Some suggested choosing one of the two countermeasures depending on the DP type. *"Combining* HL+E *and* SW *is the best.* SW *for sentences and elements that can not be clicked anyway.* HL+E *for all the rest"* (*P9*). *"A mix of* SW *for advertisements, Colors etc. [and]* HL+E *for hid-*

Several possibilities to combine *SW* and *HL+E* have been proposed

den costs would be ideal" (*P14*). Others suggest letting them work in parallel. *"Combine* HL+E *and* SW *to primarily increase the learning effect to quicker finish the ordering process"* (*P7*). *"Show me the clean version from* SW *and if a switch is toggled, show me the* HL+E *version"* (*P8*). Finally, it was suggested to use the two countermeasures after each other in a process. "SW *and* HL+E *can be combined in a way, that is a global switch that turns off the boxes (visual clutter) and another global switch that deactivates all DPs after one could get an overview with* HL+E" (*P12*).

It was recommended to combine *HD* with the explanations shown by *HL+E*

Participants also proposed combining the countermeasures *HL+E* and *HD* (4). *"A combination of* HL+E *and* HD *(but only in situations in which it adds value [...], preferably for colors but not for text) is the most suitable, I think"* (*P3*). *P17* suggested adding a *"remove such elements in the future"* option to *HL+E*.

Some other intriguing comments from individual participants were that there should be some sort of notification when a countermeasure is active (*P18*), and that users should be able to decide which elements are important and should not be changed (*P11*). Another participant thought about the countdown and came up with a method to reduce the pressure on DPs without removing any content, which was essentially *Lowlight*.

Chapter 6

Evaluation

In this chapter, we discuss our results, make some brief recommendations based on our findings, and discuss limitations.

6.1 Discussion

We first discuss our findings from Study 2, comparing them to Study 1 where possible. We summarize key findings from the qualitative data we collected and conclude with a discussion of our hypothesis.

6.1.1 Comparison of Quantitative Results

In the following, we discuss the quantitative results of Study 2 that are comparable to Study 1.

HL+E is in general the most preferred countermeasure

Rankings: The general ranking of countermeasures in Study 2 was similar to Study 1. It appears that *HL+E* is generally the preferred method. In both studies, it was ranked first most often and received the best average ranking. It is followed by *SW* as the second best method. In Study 1,

HD was very controversial. Half of its rankings were first or last (with 7 countermeasures to rank), with an almost even split. In Study 2, *HD* was ranked twice as often last as first. It seems that *HD* is less preferred in more realistic interaction scenarios.

SW was rated much better in Study 2, although it did not counteract all DPs

Ratings: In Study 2, countermeasures were rated in the same categories as in the first study, allowing for a direct comparison (Fig. 6.1). *SW* was rated notably better in Study 2. In addition, it was rated in Study 2 to be by far the clearest, while in Study 1 it was rated worse than the other two. This may be partly because we redesigned the switch icon based on feedback from Study 1, but we also think that it was harder for participants to imagine what *SW* does than to simply try it out. Most ratings for *HD* and *HL+E* were fairly constant between the two studies, but there are some exceptions. *HD* was rated less clear in Study 2. Participants were confused that some DPs in Study 2 were not counteracted by *HD*. This might explain the negative shift in the rating, but *SW* has the same issue and was rated better. In Study 2, participants rated *HL+E* to make the website feel not as good as in Study 1, which was expected since the websites in Study 2 had more DPs than just one per page, resulting in much more visual clutter. In fact, we would have expected this impact to be larger, since *HL+E* cluttered the websites in Study 2 more.

For most DPs, preferences for countermeasures were evenly split between *HD*, *HL+E*, and *SW*

Preferred countermeasures: When there is no useful information in a DP, people seem to clearly prefer *HD*. Not surprisingly, in cases where *HD* and thus *SW* could not do anything, *HL+E* was clearly preferred. What was unexpected was that in all other cases, which is more than half, the preferences were almost equally split between *HL+E*, *HD* and *SW*, even though our sample group for Study 2 was quite homogeneous. The preferred countermeasures can be compared to the first places in the DP-specific rankings from Study 1, although it should be noted that the concrete examples for the types of DPs were different between the two studies. In the ranking for *Confirmshaming*, *HL+E* was ranked first by 40%, but in Study 2, participants' preferences were evenly split between *HL+E*, *HD*, and *SW*.

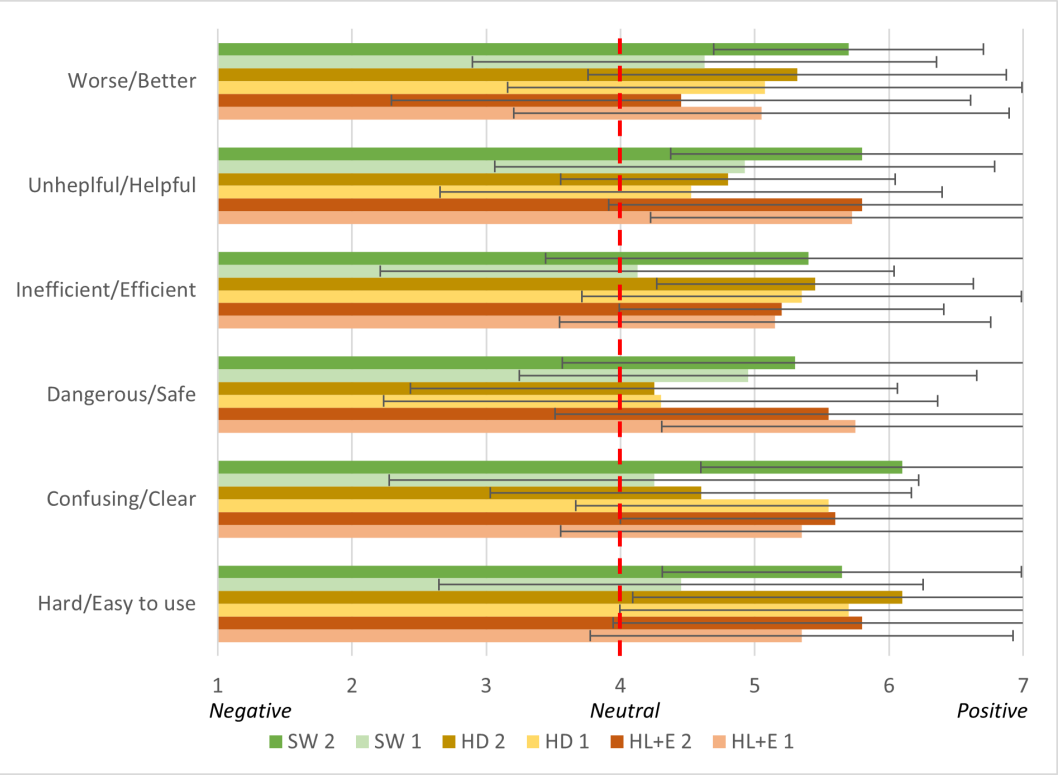


Figure 6.1: Comparison of average ratings in Study 1 and Study 2. 1 is the most negative and 7 is the most positive. The red dashed line is the neutral score of 4. The numbers behind the countermeasure names indicate the study. The black whiskers are the standard deviation. *SW* was rated much better in Study 2, while for *HL+E* and *HD* there were no big differences in most cases.

On the one hand, this could be explained by the fact that *Confirmshaming* was more aggressive in Study 2. On the other hand, some participants seem to prefer the tidy feel of *HD* when experiencing real interaction, and when they still wanted to see the original content, they chose *SW*. In the *Low Stock Message* ranking in Study 1, 37.5% of the participants ranked *HL+E* first, but 40% ranked *HD* first. In Study 2, the *Low Stock Message* was not part of the study, but we expect it to behave similarly to the *Activity Message* (A) and the *High Demand Message* (D). They look similar and use similar mechanisms to pressure the customer. In both cases, *HD* was clearly preferred (60% and 75%). Participants deemed them to be not useful and irrelevant. We think that in Study 2, participants liked *HL+E* less because

these DPs try to grab the user's attention and *HL+E* only supports that. This seemed to be especially noticeable to participants when solving a task. For visual interference, the situation is very similar. In Study 1, *HL+E* was ranked first with 32.5% and *HD* with 37.5%. But in Study 2, *HD* was clearly preferred with 65%. Participants found the visual interference in Study 2 to be misleading, perhaps because it was more aggressive than in Study 1.

6.1.2 Study 2 Specific Quantitative Results

Additional measurements could be made that were not possible in Study 1 because Study 2 included real interaction. We discuss them in this section.

The positive impact of *HD* on usability is smaller than expected

Usability: In most cases, the *Ticket Shop* was less usable than the *Smartphone Shop* (Fig. 5.4), but surprisingly the opposite was true for *HL+E*. It seems that *HL+E* works particularly poorly on the *Smartphone Shop* in terms of usability. Overall, *HD* and *SW* slightly improved the usability compared to the baseline, while *HL+E* did not. We would have expected *HD* to have a larger positive impact because it removes unnecessary content, clarifies wording, and removes misleading coloring.

Countermeasures seem to simplify interaction with malicious websites

Efficiency: It seems that the countermeasures make interaction easier, especially on the more difficult to use *Ticket Website* (Fig. 5.6). Participants were also less confident on the *Ticket Website*. *HL+E* and *SW* increased confidence for this page, *HD* did not. Users seem to take longer to reach their goal when using countermeasures. For *SW*, the difference from the baseline was particularly large. This may be due to a limitation in our prototype that forced participants to deactivate one switch before activating another. Even without this limitation, *SW* introduces the most additional interaction and is expected to take the longest. Surprisingly, *HD* did not reduce the average task completion time.

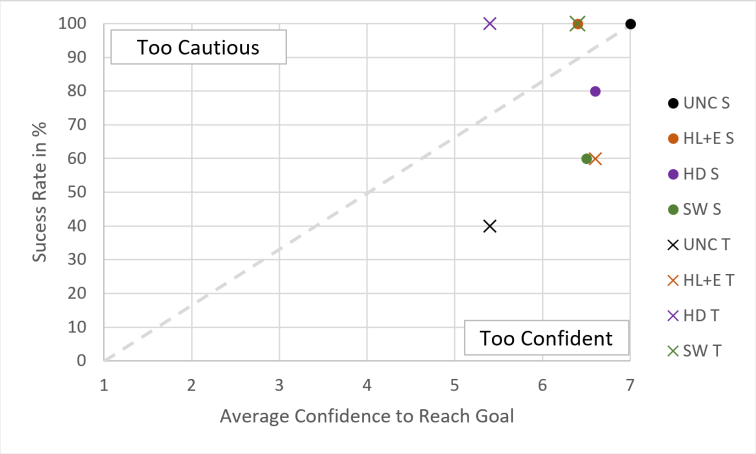


Figure 6.2: The results of success rate and confidence are plotted on a scatter plot. The different countermeasures are color coded. The dots represent the results for the *Smartphone Shop* and crosses the results for the *Ticket Shop*. The combination of countermeasures and websites seems to have a complicated effect on overconfidence. More data may be needed to identify a pattern. (n=5)

It is ambiguous whether the countermeasures increase the success rate

Not getting fooled by DPs: Preventing users from being tricked by DPs is the main goal of countermeasures. On average, the success rate was slightly increased by the countermeasures (Tab. 5.4). Surprisingly, *HD* and especially *SW* decreased the success rate for the *Smartphone Shop*, which was used error-free in the baseline. Countermeasures that create a false sense of security are particularly undesirable. Figure 6.2 compares participants' confidence with the success rate. *HD* made participants the least overconfident for both website prototypes.

HL+E makes pages feel less trustworthy, *HD* and *SW* increase trust

Perceived trustworthiness: Our two website prototypes contained a high amount of DPs, and thus were both perceived as slightly sketchy (as opposed to trustworthy). As we expected, *HL+E* seems to decrease the perceived trustworthiness of a page, while *HL+E* increases it. However, we were surprised to find that *SW* also increases a page's trustworthiness almost as much as *HD*. This was unexpected, since *SW* exposes the manipulative intent of ser-

vice providers by directly comparing the DP to a non-manipulative version. It seems that the hiding effect *SW* has by showing the non-manipulative version by default is already sufficient for users to make the page feel more trustworthy. *HD* and *SW* do not work for all DP types. So it may be problematic that they make websites feel more trustworthy.

6.1.3 Comparison of Qualitative Results

The comments on specific countermeasures were similar to Study 1, although we had mostly different participants. Some additional comments, primarily related to interaction, only occurred in Study 2. Below, we summarize and discuss the qualitative feedback by countermeasure.

Qualitative Discussion of *HL+E*:

HL+E adds a lot of visual clutter, redesigning may help reduce this

Influence on page perception: In Study 1, many participants found *HL+E* too visually cluttered. They expected this to be particularly strong on pages containing multiple DPs at once. This setting was tested in Study 2 and confirmed by comments in that study. Participants mentioned that this method could make pages feel less trustworthy. But especially in Study 2, it was also commented that this would only affect pages that deserve it. We think that the possibility of being fooled by the DPs, which was only present in Study 2, made the participants appreciate the warnings more. It seems feasible to investigate a redesign of the visual appearance of *HL+E*, participants complained about the visual design.

In some cases, drawing attention to a DP may be undesirable

Directing attention: After being able to interact with our prototypes, some participants commented that they would not have noticed certain DPs, such as the *Activity Message*, if they were not highlighted prominently. This emphasizes the characteristic of *HL+E* to draw users' attention to the

DPs. This can be counterproductive for DPs that are designed to grab attention and induce pressure.

Using the explanations: In Study 1, one concern expressed by participants was that the explanations provided by this countermeasure would not be read because they had to be actively accessed. In Study 2, some participants commented that the explanations were cumbersome, but it seems that this was mainly referring to the task of reading the explanations themselves, rather than accessing them. We did not observe any participants in Study 2 who did not want to read the explanations. Participants in both studies found this method helpful in increasing their knowledge and awareness of DPs. They suggest options to adjust the behavior of *HL+E* to be less annoying in the long run, for example by allowing to disable certain warnings once the users feel like they do not need them.

Accessing the explanations is easy, reading them can be tedious

Keeping original content: In Study 2, participants commented that *HL+E* did not hide or alter potentially relevant content on the page. This was considered positive. It seems that this advantage is particularly apparent after interacting with *SW* and *HD*, as the two participants who commented on this both had *HL+E* as their last condition.

Some participants liked that *HL+E* does not change original content

Qualitative Discussion of *HD*:

Removing manipulation: *HD* was seen as having the potential to completely stop the manipulation by DPs if it works flawlessly. However, in Study 2, participants did not like the fact that not all types of DPs could be countered. They remarked that users might let their guard down when using this method, and thus be hit even harder in such cases. Participants were also uncomfortable that sketchy sites could appear deceptively trustworthy when the DPs are removed.

Participants also saw potential drawbacks to removing DPs

Participants liked that *HD* reduces cognitive load

Reducing cognitive load: In both studies, participants commented that *HD* can improve interaction with websites by making them clearer or reducing the cognitive load on the user. However, this comes at the cost of potentially decreasing user awareness of DPs over time.

HD would have to work flawlessly, since changes to the page are not indicated

Trust in the method: This method requires user trust. In both studies, many participants commented that it felt unsafe or dangerous because their content was changed automatically and silently. In Study 1, a version of *HD* was included that marked changed areas (*HD+M*). Participants gave strong negative feedback for this approach, as it is confusing when hidden information is highlighted. *SW* may be a compromise for that.

Qualitative Discussion of *SW*:

The ability to revert to the original element allows users to build confidence in the method

Access to the original: In both studies, participants liked that *SW* allows users to return to the original view if they wish. This makes it possible to build confidence in the technology over time and to double-check when a generated DP alternative looks wrong. Since the cleaned version of a DP is displayed by default, users can decide how much DP manipulation they want to be exposed to.

Users could be distracted by *SW*

Efficiency: The main criticism of *SW* was that it would be distracting in everyday use. It could reduce efficiency when interacting with a website. This is supported by the longer task completion times we measured for *SW*. In Study 2, some participants found the individual switches cumbersome. It was suggested to introduce a global switch, which would toggles all switches.

6.1.4 Hypothesis

In the following, we discuss the tendencies our results show regarding our hypotheses (Section 5.1). A signifi-

cance analysis is not included, we decided to leave it out of the scope for this thesis.

H1.1: *Different countermeasures have different usability.*
Probably yes. We observed different average SUS scores for the countermeasures (Tab. 5.2). Surprisingly, *HL+E* strongly decreased the usability of the *Smartphone Shop*, while it had no big impact on the *Ticket Shop*.

H1.2: *Some countermeasures will enhance the usability of a page.*
Probably yes. *HD* and *SW* improved the average SUS scores of both website prototypes in similarly. *HL+E* did not.

HD and *SW* tend to improve usability

H2.1: *Different countermeasures will affect the efficiency of the user.*
Perhaps. For this we consider success rate, task completion time, perceived ease of achieving the goal, and confidence in having achieved the goal. No countermeasure clearly improved the success rate for both website prototypes (Tab. 5.4), but for the *Ticket Shop* it was improved by all countermeasures. Participants who used *SW* took on average considerably longer to complete the tasks (Fig. 5.7). Surprisingly, none of the countermeasures accelerated the participants in completing their tasks, not even *HD*. Countermeasures tended to make it easier for participants to achieve their goals on the websites, especially on the *Ticket Shop*. It seems that participants' confidence was generally high for *Smartphone Shop*. While for *Ticket Shop* it was increased by *HL+E* and *SW*, but not by *HD* (Fig. 5.6).

Countermeasures did not make participants faster at solving the tasks

H2.2: *Some countermeasures will enhance the efficiency of the user.*
Not in general, but probably for some websites. No countermeasure increased the speed of task completion (Fig. 5.7), and no countermeasure consistently improved the success rate on both website prototypes (Tab. 5.4). However, on the *Ticket Shop HD* and *SW* increased the success rate, *HL+E* and *SW* increased the confidence, and all countermeasures increased the easiness (Fig. 5.6).

Countermeasures improved efficiency on some websites, but not on others

H3.1: *Certain countermeasures will be preferred in general.*
Probably yes. The rankings of Studies 1 and 2 indicate that

HL+E is generally preferred

HL+E is preferred by the most users. In Study 2, *SW* also scored well in most categories, much better than in Study 1.

For most DPs, no countermeasure was clearly preferred

H3.2: *Certain countermeasures will be preferred for certain DPs.* Partially yes. The preferred countermeasures for each DP in Study 2 show strong differences for some DPs, but for more than half of the DPs, preferences are fairly evenly split between *HL+E*, *HD*, and *SW* (Fig. 5.10). *HL+E* is preferred over *UC* for elements that cannot be overridden by *HD* and *SW*. *HD* is preferred for *Activity Message*, *High Demand Message* and *Visual Interference*.

HL+E reduces perceived trustworthiness

H4: *Countermeasures will impact how trustworthy a page feels.* Probably yes. The website prototypes on average felt less trustworthy for *HL+E* and more trustworthy for *HD* and *SW* (Fig. 5.5). This is unexpected for *SW*, as this countermeasure reveals how the website would have tried to manipulate users.

6.2 Recommendations for DP Research

In this section, we will briefly discuss some suggestions and considerations we have regarding our methodology for evaluating and developing countermeasures.

6.2.1 Study Setups

Online and lab study had fairly consistent results

Our approach was to first conduct an online screenshot-based study, and later perform a more in-depth interaction-based lab study. This yielded more consistent results than we expected. Our data suggests that, especially for countermeasures that are not interaction-heavy, a screenshot-based study provides a good approximation to more complicated study setups that require interactive prototypes. This makes us optimistic that future research on the methods we exclusively investigated in Study 1 (*HL*, *HD+M*, *LL*) supports our findings.

6.2.2 Quality Criteria for Countermeasures

Different environments may require different quality criteria for countermeasures

Based on the objective, the quality of a countermeasure can be measured by several metrics. In our studies, we investigated the online shopping context. We believe that in this context, the reduction of DP induced user errors should be the top priority, as mistakes can be costly. In other, more casual situations, usability might be the top priority. In a business setting, task completion time could be prioritized. Other relevant characteristics are how confident users feel when using the countermeasure, how the method helps users correctly assess whether they can trust a website, whether the countermeasure increases user awareness, and whether it is aesthetically pleasing.

6.2.3 Suggestions for Developers

For creators of DP countermeasures, we want to provide some preliminary suggestions. They are not yet supported by large samples and a significance analysis.

- Select *HL+E* if a single countermeasure approach must be picked.
- We have not investigated combined approaches, but it was frequently suggested to combine *HL+E* and *SW* in a variety of ways.
- DPs that are not interactive and do not provide any concrete information to the user other than pressuring them, such as the typical *High-Demand Messages*, can be deleted.
- Remove *Visual Interference* if possible.
- Apply *HL+E* to DPs that try to be sneaky, like *Hidden Subscription*.
- *HD* has a lot of potential, but it needs to prove that users can trust it to work properly. Be aware that not all DPs are removable on client-side.

- User preferences are remarkably heterogeneous, so make the countermeasure customizable.
- Include a simple *off* switch that deactivates the countermeasure on the current page.

6.3 Limitations

Responses from participants were diverse, a larger sample size would be useful to make more reliable statements

Study 2 was designed to investigate how users perceive three different countermeasures when interacting with a website. We created interactive prototypes that were not actual websites. We compromised between the complexity of the page and the number of DPs used in order to keep the conditions comparable, literature-based, and practical. Most of our participants had a background in technology, which made our sample rather uniform. This is unlikely to represent the full population of web users. We had 20 participants in Study 2. This allowed us to iterate our counterbalancing plan two and a half times. However, the nature of our setup partitioned the number of data points for some measures to 10 or 5. This allowed us to avoid learning effects where we expected them to be strong, but also reduced our sample size for those situations. We recommend increasing the sample size. For the sake of transparency, the number of data points is reported individually for all measures. We decided to not include a significance analysis in this thesis due to its scale. Although we cannot predict how our results will translate to long-term interaction in an everyday setting, we attempted to reduce novelty effects by providing a practice page for each condition before the actual tests began. In both studies, we compared different countermeasures to see what users prefer in certain situations. We gave them different options to choose from. This may influence the responses because all options are known to the participant in advance and there is no risk. This might bias the results for countermeasures that fail big, like *HD*. We think this effect was reduced in our general ranking questions, since there were no examples given.

Chapter 7

Summary and Future Work

7.1 Summary

In this thesis, we investigated from an HCI perspective how Dark Patterns (DPs) could be visually counteracted to minimize their manipulative effect on website users. We first presented the state of research on DPs, how they can be countered in general, and what technologies are being developed to automatically detect them. We provided context for the idea of visual DP countermeasures by discussing four different leverages that can be used against DPs: *Laws*, *Public Pressure*, *Awareness of Users* and *Technically Weaken DPs*. Based on literature and own ideas, we presented a set of ideas to counteract DPs with systems such as browser add-ons. To evaluate them, we conducted two studies that examined how users perceive different countermeasures. The first study compared several countermeasures in an online study using screenshot prototypes. Based on Study 1, we designed and conducted a second study that examined a smaller subset of three countermeasures using two interactive shopping website prototypes. The prototypes included nearly all DP types from an established taxonomy. Because Study 2 included real-world interactions, we were able to look more closely at the impact of countermeasures

We developed several ideas for DP countermeasures and investigated them in two studies

on efficiency, usability, trust, and success rates. In addition, we examined which countermeasures are preferred in which situations.

We have found, that users like countermeasures and clearly prefer them. In general, the method *Highlight with Explanation* is preferred. It marks manipulative elements and provides explanations, but has the disadvantage of visually cluttering the site. Removing the explanations could reduce this, but was not well received. *Hide without Marking*, a method that simply removes the DPs and replaces them with non-manipulative elements, was considered highly controversial. It makes pages clearer and improves usability, but it also makes users fear that they will be deprived of relevant content in the event of an error. It could also make sketchy websites appear misleadingly trustworthy. Marking cleaned areas to make the method more transparent has not been well received. *Switch* is a compromise between hiding the DPs and being transparent. It was ranked second best, but has a negative impact on efficiency. There appears to be no single method that solves all problems; preferences vary widely among users and should depend on the type of DP they are dealing with, as well as the general context. More research is needed, but we offer some preliminary suggestions for developers of DP countermeasures in Section 6.2.3.

Different methods have different strengths and weaknesses, they should be selected depending on the context

7.2 Future Work

We present our ideas for future research topics, gradually moving from areas close to our studies to broader topics.

The participant groups, especially in Study 2, were quite homogeneous. By conducting this study with participants from various backgrounds, it would be possible to compare the preferences of different user groups.

Study a broader population

In Study 2, there were situations where the success rate of the participants decreased compared to the baseline. This should be investigated further. Perhaps our sample was not large enough, or perhaps users became overconfident

Examine how to maximize the success rate

and relied too much on the countermeasures. This is especially relevant since we cannot expect real-world implementations of countermeasures to work flawlessly.

Explore visual redesigns of the countermeasures

Participants commented on the visual design of the countermeasures throughout the studies. For example, the red boxes and warning signs of *Highlight with Explanation* were perceived as threatening. Studying the influence of different visual designs on the perception of DPs would help in choosing how to design the appearance of countermeasures. Some participants suggested that countermeasures and their visual design could influence the goals that users pursue. For example, users might initially be interested in faster shipping, but an alarming red box might cause them to change their goal. This would go against our intentions for countermeasures, as it would introduce a new form of user manipulation.

Investigate the influence of DP aggressiveness on user preferences

We tested a single representation for each DP, however DPs like *Visual Interference* can appear in many appearances and levels of aggressiveness. It would be promising to compare which countermeasures participants prefer depending not only on the type of DP, but also on its aggressiveness.

Study user preferences for non-shopping websites

We focused on shopping websites for desktop. Other usage scenarios may affect user preferences, since on a shopping website real money is at stake, while on a social media platform it might be privacy options. It can also be assumed that users' priorities change in a mobile context, as there is less screen space available, the mode of input changes, and users are more likely to be distracted by the world around them. In addition, the DPs that users encounter on mobile devices are different [Gunawan et al., 2021].

Examine effects of countermeasure failures on interaction

Imperfect implementations can also have a huge impact on how users perceive countermeasures like *Hide without Marking*. Methods for correcting errors in a community-based approach, similar to ad blockers, might reduce this, but it also seems relevant to investigate for different countermeasures how error prove they need to be for users to accept them. We expect users to have a higher error tolerance for methods that do not change the content of the page.

Use logging and process analysis to investigate interaction

It may be interesting to use process analysis approaches to find out how users embed countermeasures in their website interactions. By using a platform other than Figma for the prototypes, it would be possible to accurately log user behavior and extract how participants use the switch functionality or tooltips. Process analysis tools could be used to determine how much additional information should be provided by the countermeasure and when it could be omitted to reduce visual clutter.

Use countermeasure characteristics as factors in a large factorial study

A large study using a factorial approach would provide valuable insights on how different characteristics of countermeasures play together. Factors could be marking of DPs (Highlighting/Lowlighting/Nothing), Information (Name/Name+Explanation/Nothing), showing a non-manipulative alternative (Yes/No) and local off switch (Yes/No).

Investigate our other countermeasure ideas

In the future, it would be interesting to investigate the countermeasures that we identified in Chapter 3 which we were not able to include in our study. We expect that additional features, such as DP warnings next to search engine results, will be needed to develop holistic solutions.

To summarize, we have identified promising additional research questions:

- How can countermeasures be integrated most intuitively and efficiently into users' natural workflow?
- What quality criteria of a countermeasure are prioritized by users in different contexts?
- How does user perception of the countermeasure change over longer times of use?
- What should countermeasures look like in different contexts?
- How aggressive should countermeasures be for different DPs?
- How do users want to customize countermeasures if they have access to settings?

Finally, the biggest and most important step is to continue developing the technical foundation for DP countermeasures so that our findings can eventually be applied in the real world to help people in their daily lives.

Implement DP countermeasures

Appendix A

Study 1 Prototypes

In the following, we present the screenshot prototypes used in Study 1.

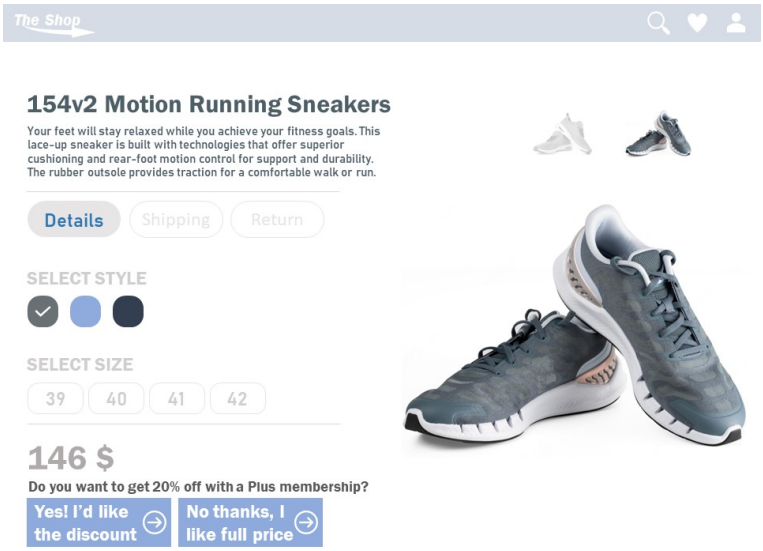


Figure A.1: Screenshot prototype using *Confirmshaming* without any countermeasure.

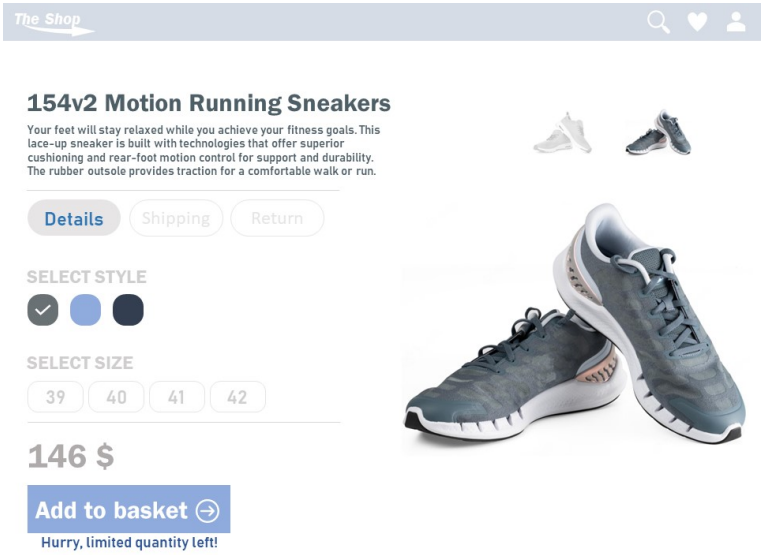


Figure A.2: Screenshot prototype using *Low Stock Message* without any countermeasure.

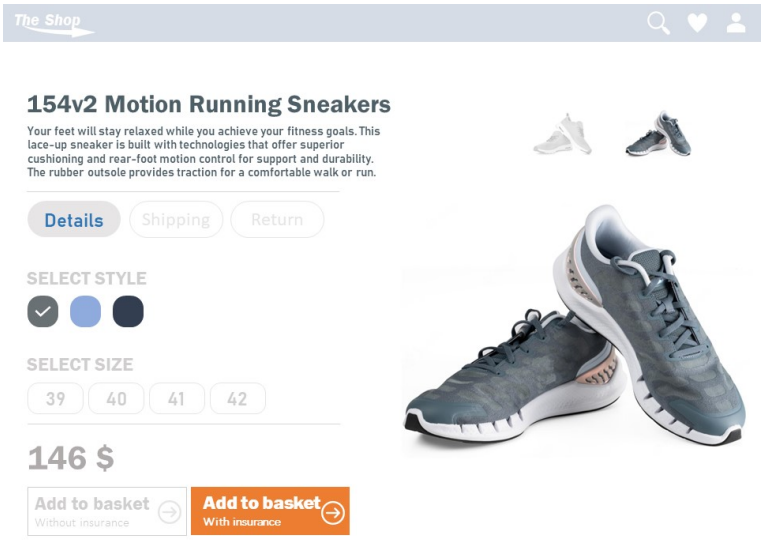


Figure A.3: Screenshot prototype using *Visual Interference* without any countermeasure.

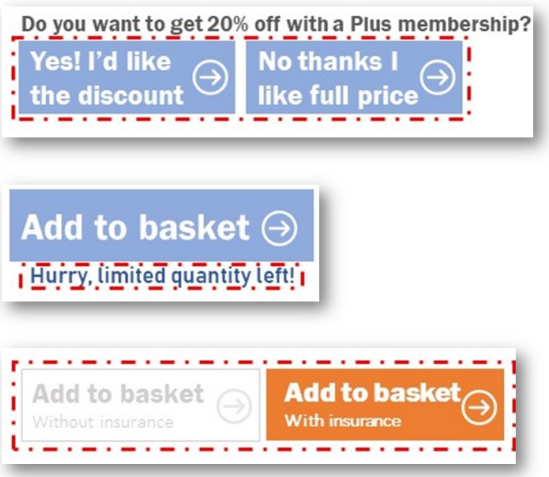


Figure A.4: Collection of *Highlight without Explanation* applied to the three DPs.

Figure A.5: Collection of *Highlight with Explanation* applied to the three DPs.

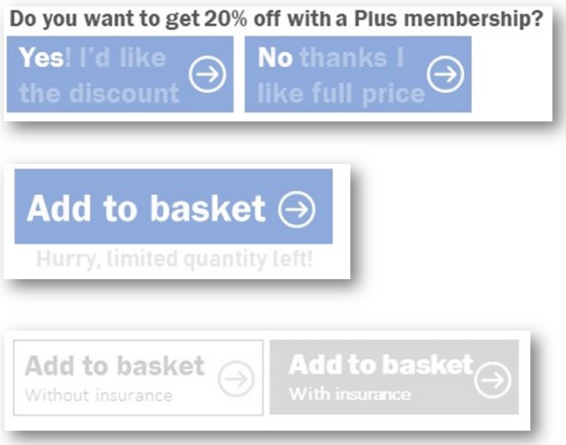


Figure A.6: Collection of *Lowlight* applied to the three DPs.

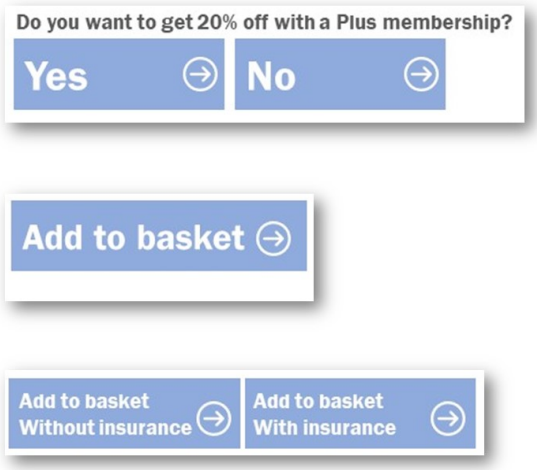


Figure A.7: Collection of *Hide without Marking* applied to the three DPs.

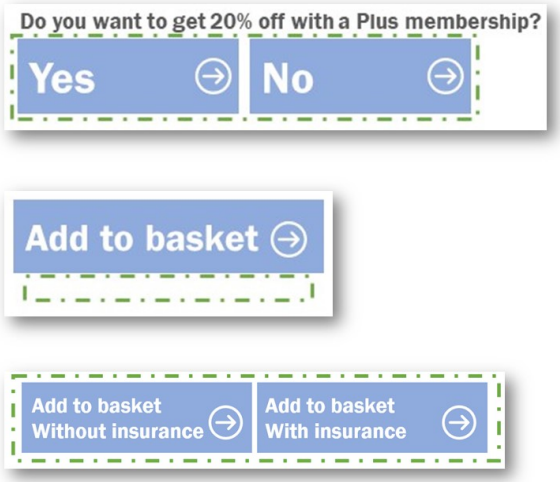


Figure A.8: Collection of *Hide with Marking* applied to the three DPs.

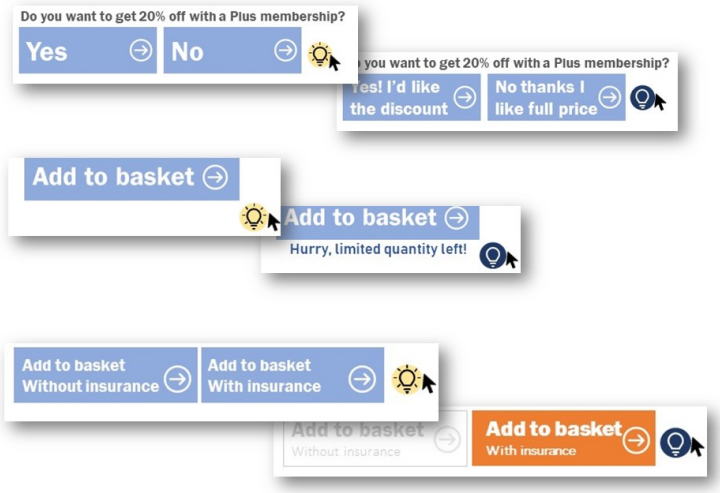


Figure A.9: Collection of *Switch* applied to the three DPs.

Appendix B

Study 2 Questionnaire

In the following, we present the questionnaire of Study 2. The two pages with questions referring to the *method* were included four times, once for each condition. We also created a German version of the questionnaire.

ID

Questions at the beginning of the study:

Demographie:

1. Age: ____
2. Gender: _____
3. Profession: _____
4. Field of work: _____

Dark Patterns:

Dark patterns are elements of user interfaces that aim to influence a person's behaviour against their intentions or interests. There are many different types of dark patterns. Often they are used by companies to make more money or get more user data.

5. How aware were you of the existence of dark patterns on websites?
   Not aware at all ◯ ◯ ◯ ◯ ◯ ◯ ◯ *Very aware*

Other studies:

6. Have you already participated in a study on Dark Patterns?
   ◯ Yes
   ◯ No

7. Have you already participated in our online study on "Dark Pattern Countermeasures"?
   ◯ Yes
   ◯ No

ID

M _____ T_____ C_____

Goal:

- To achieve the aforementioned goal was...
  Very difficult ◯ ◯ ◯ ◯ ◯ ◯ ◯ *Very light*
- How sure are you that you have achieved the aforementioned goal?
  Not confident at all ◯ ◯ ◯ ◯ ◯ ◯ ◯ *Very confident*
- Comments

Website:

These questions refer to the overall impression of the website the way you have just used it.

- This website seemed to me...
  Very dubious ◯ ◯ ◯ ◯ ◯ ◯ ◯ *Very trustworthy*

Please tick one answer per question, depending on how far you intuitively agree with the statement.

	Strongly disagree				Strongly agree
1. I think I would like to use this website frequently.	◯	◯	◯	◯	◯
2. I found the website unnecessarily complex.	◯	◯	◯	◯	◯
3. I thought the system was easy to use.	◯	◯	◯	◯	◯
4. I think I would need the support of a technical person to be able to use this site.	◯	◯	◯	◯	◯
5. I found that the various functions are well integrated into the website.	◯	◯	◯	◯	◯
6. I thought that there was too much inconsistency in this website.	◯	◯	◯	◯	◯
7. I would imagine that most people would learn to use this system very quickly	◯	◯	◯	◯	◯
8. I found it awkward to use the website.	◯	◯	◯	◯	◯
9. I felt very confident using the website.	◯	◯	◯	◯	◯
10. I needed to learn a lot of things before I could get going with this system.	◯	◯	◯	◯	◯

ID

Method:

*These questions relate to the **Method**, which has just been applied to the website, not to the Website itself.*

- I think for me as a user this method would be in practice
  1. *Confusing* ◯ ◯ ◯ ◯ ◯ ◯ ◯ *Clear*
  2. *Unhelpful* ◯ ◯ ◯ ◯ ◯ ◯ ◯ *Helpful*
  3. *Dangerous* ◯ ◯ ◯ ◯ ◯ ◯ ◯ *Safe*
  4. *Hard to use* ◯ ◯ ◯ ◯ ◯ ◯ ◯ *Easy to use*
  5. *Inefficient* ◯ ◯ ◯ ◯ ◯ ◯ ◯ *Efficient*

- I think when this method is used, most websites feel
  1. *Worse* ◯ ◯ ◯ ◯ ◯ ◯ ◯ *Better*

Further comments regarding the method or the website:

ID

Questions at the end of studie:

You have now seen these four methods:

- **Unchanged (UNC):** The manipulative element is not altered in any way.
- **Box (BOX):** A red dashed box and a warning sign are drawn around the manipulative element. If the mouse pointer hovers over it, an explanation of why it is selected is displayed.
- **Removed (REM):** If the manipulative element can be visually removed or rephrased, this is done. Otherwise, the original will be shown.
- **Switch (SWT):** If the manipulative element can be visually removed or rephrased, this is done. In such cases, a switch next to the element allows you to switch to the original.

1. Please place the methods according to how much you would actually like to use them and justify your choice:

Rank	**Method**
Best Place	_____
2nd Place	_____
3rd Place	_____
Last Place	_____

Justification:

ID

2. For each highlighted item, please indicate which method you would prefer:

Should build up pressure to buy the product

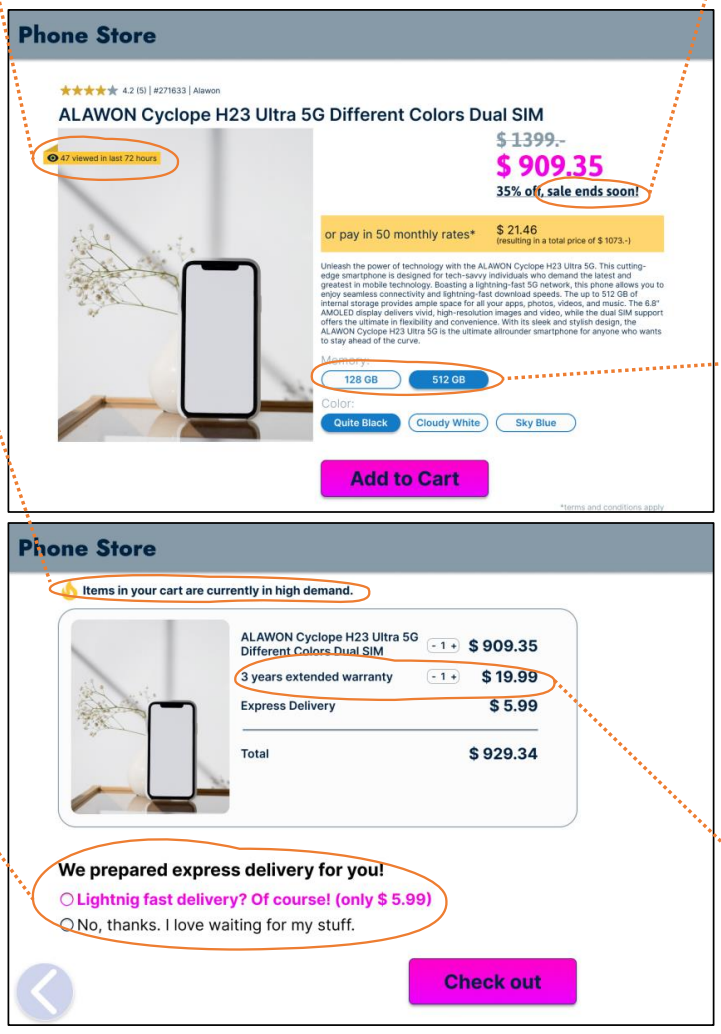
A) _____

Justification:

End time open to increase pressure

B) _____

Justification:



Should build pressure to buy quickly

D) _____

Justification:

More expensive variant is pre-selected

C) _____

Justification:

Manipulatively designed and formulated

F) _____

Justification:

Automatically added to basket, must be actively removed

E) _____

Justification:

ID

Quotes of questionable authenticity

G) _____

Justification:

Countdown to increase pressure

H) _____

Justification:

Expensive option is strongly highlighted

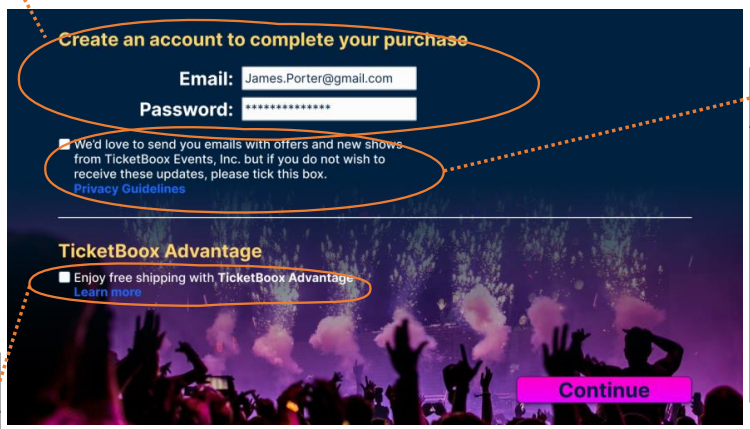
I) _____

Justification:

Data collection by forced account

J) _____

Justification:



Misleading wording. Tick = No newsletter

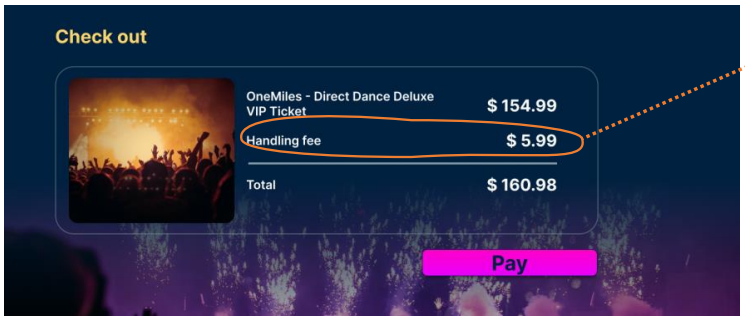
K) _____

Justification:

Hidden subscription (info with additional click)

L) _____

Justification:



Hidden additional costs late in the process

M) _____

Justification:

ID

4. Please describe the strengths and weaknesses of each method:

Method	Strengths	Weaknesses
Unchanged		
Box		
Remove		
Switch		

5. Do you have suggestions for improvement of the methods or ideas for further methods?

Appendix C

Study 2 Prototypes

This appendix contains pictures of our prototypes for Study 2. A printout of these was also provided to the participants as a reference when they decided which countermeasure they preferred for which DP at the end of the study.

Muffin Deluxe

Deluxe Muffins with Extra Chocolate Chips

Indulge in the classic flavor of freshly baked chocolate chip muffins. Made with rich and velvety chocolate chips nestled in a soft and tender muffin base, these sweet treats are the perfect snack or breakfast option. Moist, fluffy, and bursting with chocolate in every bite, they're sure to satisfy your sweet tooth cravings.

Select Quantity:

4 Pieces | 10 Pieces | 24 Pieces

Price:

0.00 $

Buy

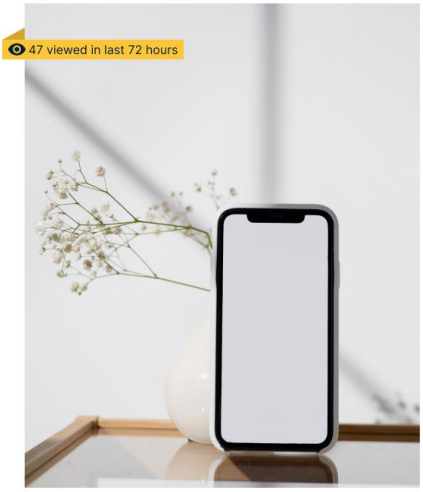
Hurry! Only few left!

Unchanged (UNC)

Phone Store

★★★★★ 4.2 (5) | #271633 | Alawon

ALAWON Cyclope H23 Ultra 5G Different Colors Dual SIM



~~$ 1399.-~~

$ 909.35

35% off, sale ends soon!

or pay in 50 monthly rates* $ 21.46 (resulting in a total price of $ 1073.-)

Unleash the power of technology with the ALAWON Cyclope H23 Ultra 5G. This cutting-edge smartphone is designed for tech-savvy individuals who demand the latest and greatest in mobile technology. Boasting a lightning-fast 5G network, this phone allows you to enjoy seamless connectivity and lightning-fast download speeds. The up to 512 GB of internal storage provides ample space for all your apps, photos, videos, and music. The 6.8" AMOLED display delivers vivid, high-resolution images and video, while the dual SIM support offers the ultimate in flexibility and convenience. With its sleek and stylish design, the ALAWON Cyclope H23 Ultra 5G is the ultimate allrounder smartphone for anyone who wants to stay ahead of the curve.

Memory:

128 GB | **512 GB**

Color:

Quite Black | Cloudy White | Sky Blue

Add to Cart

*terms and conditions apply

Phone Store

🔥 **Items in your cart are currently in high demand.**

ALAWON Cyclope H23 Ultra 5G Different Colors Dual SIM	- 1 +	$ 909.35
3 years extended warranty	- 1 +	$ 19.99
Express Delivery		$ 5.99
Total		$ 929.34

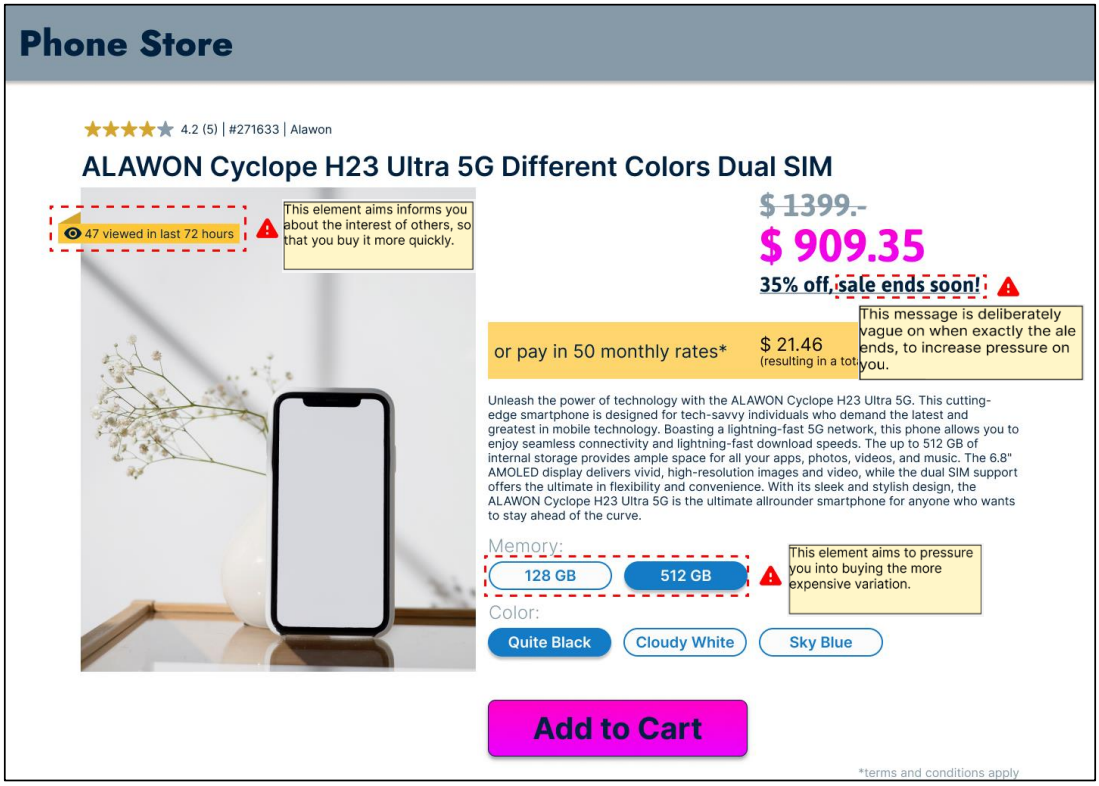
We prepared express delivery for you!

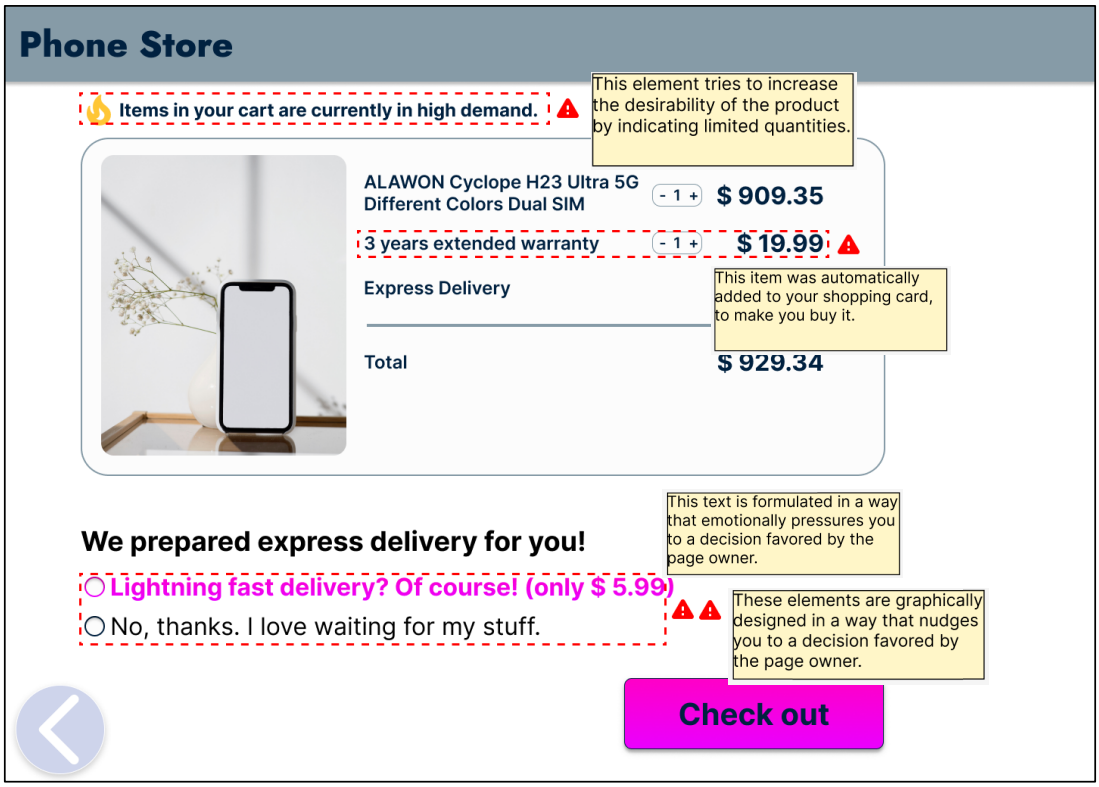
◯ **Lightnig fast delivery? Of course! (only $ 5.99)**

◯ No, thanks. I love waiting for my stuff.

Check out

Highlight with Explanation (HL+E)





Hide (HD)

Phone Store

★★★★★ 4.2 (5) | #271633 | Alawon

ALAWON Cyclope H23 Ultra 5G Different Colors Dual SIM

~~$ 1399.-~~

$ 909.35

35% off

or pay in 50 monthly rates*	$ 21.46 (resulting in a total price of $ 1073.-)

Unleash the power of technology with the ALAWON Cyclope H23 Ultra 5G. This cutting-edge smartphone is designed for tech-savvy individuals who demand the latest and greatest in mobile technology. Boasting a lightning-fast 5G network, this phone allows you to enjoy seamless connectivity and lightning-fast download speeds. The up to 512 GB of internal storage provides ample space for all your apps, photos, videos, and music. The 6.8" AMOLED display delivers vivid, high-resolution images and video, while the dual SIM support offers the ultimate in flexibility and convenience. With its sleek and stylish design, the ALAWON Cyclope H23 Ultra 5G is the ultimate allrounder smartphone for anyone who wants to stay ahead of the curve.

Memory:

128 GB | 512 GB

Color:

Quite Black | Cloudy White | Sky Blue

Add to Cart

*terms and conditions apply

Phone Store

ALAWON Cyclope H23 Ultra 5G Different Colors Dual SIM	- 1 +	$ 909.35
3 years extended warranty	- 1 +	$ 19.99
Express Delivery		$ 5.99
Total		$ 929.34

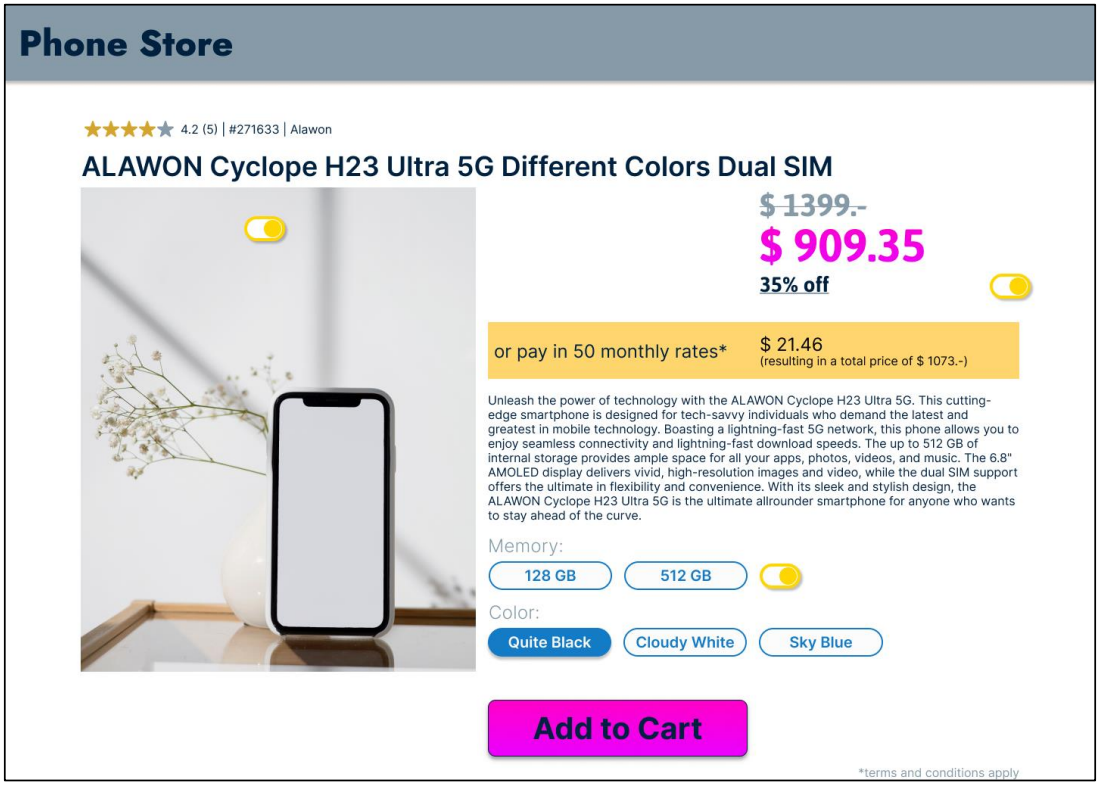
Express delivery?

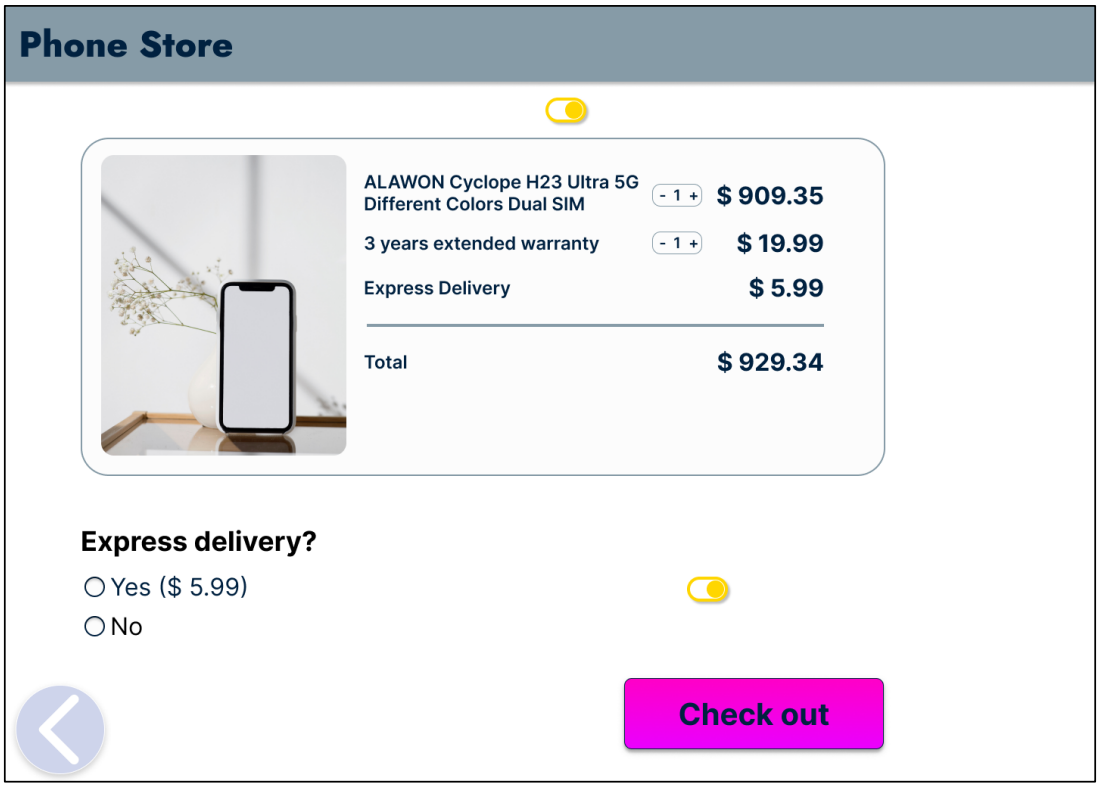
○ Yes ($ 5.99)

○ No

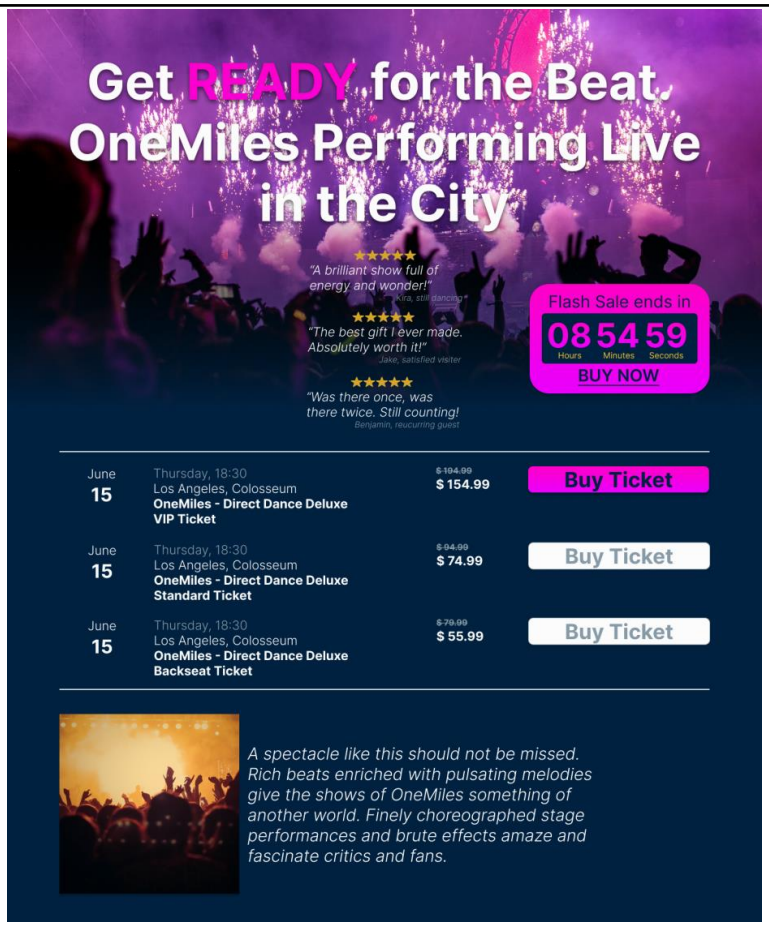
Check out

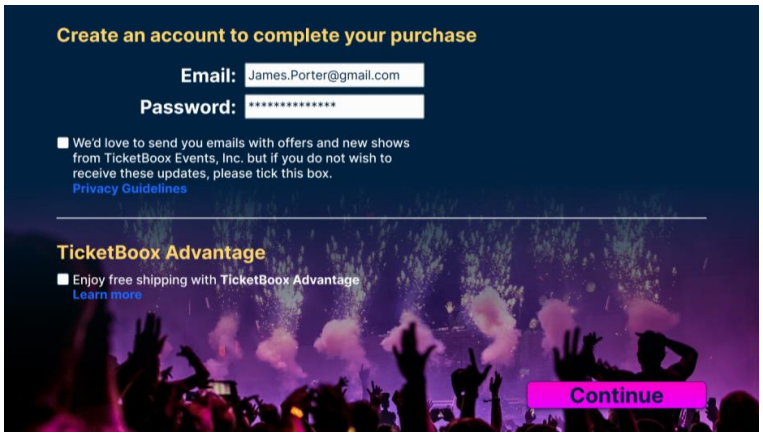
Switch (SW)

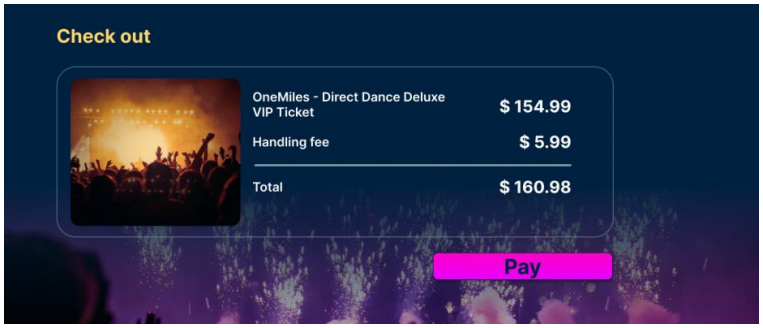

Phone Store
4.2 (5) | #271633 | Alawon
ALAWON Cyclope H23 Ultra 5G Different Colors Dual SIM
$ 1399.-
$ 909.35
35% off
or pay in 50 monthly rates*
$ 21.46
(resulting in a total price of $ 1073.-)
Unleash the power of technology with the ALAWON Cyclope H23 Ultra 5G. This cutting-edge smartphone is designed for tech-savvy individuals who demand the latest and greatest in mobile technology. Boasting a lightning-fast 5G network, this phone allows you to enjoy seamless connectivity and lightning-fast download speeds. The up to 512 GB of internal storage provides ample space for all your apps, photos, videos, and music. The 6.8" AMOLED display delivers vivid, high-resolution images and video, while the dual SIM support offers the ultimate in flexibility and convenience. With its sleek and stylish design, the ALAWON Cyclope H23 Ultra 5G is the ultimate allrounder smartphone for anyone who wants to stay ahead of the curve.
Memory:
128 GB
512 GB
Color:
Quite Black
Cloudy White
Sky Blue
Add to Cart
*terms and conditions apply


Phone Store
ALAWON Cyclope H23 Ultra 5G Different Colors Dual SIM
- 1 +
$ 909.35
3 years extended warranty
- 1 +
$ 19.99
Express Delivery
$ 5.99
Total
$ 929.34
Express delivery?
Yes ($ 5.99)
No
Check out

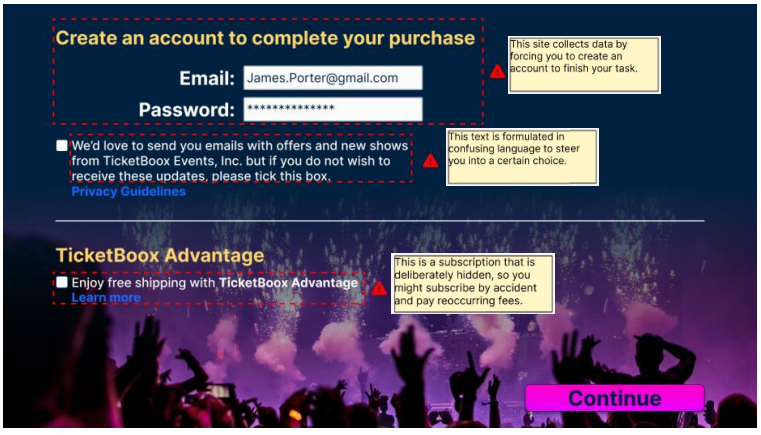
Unchanged (UNC)

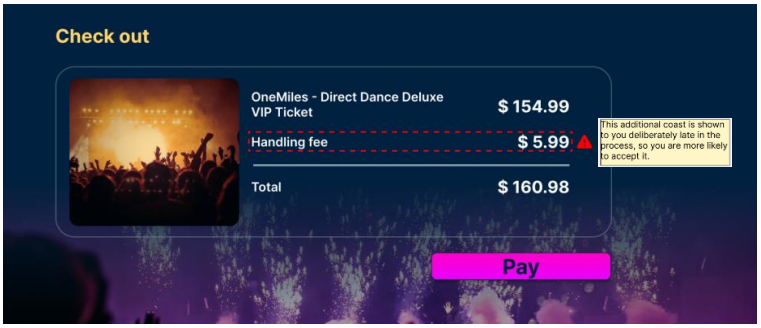






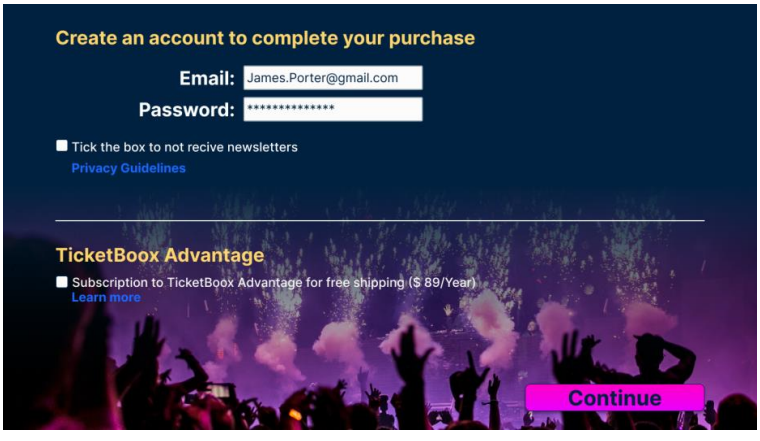
Highlighting with Explanation (HL+E)

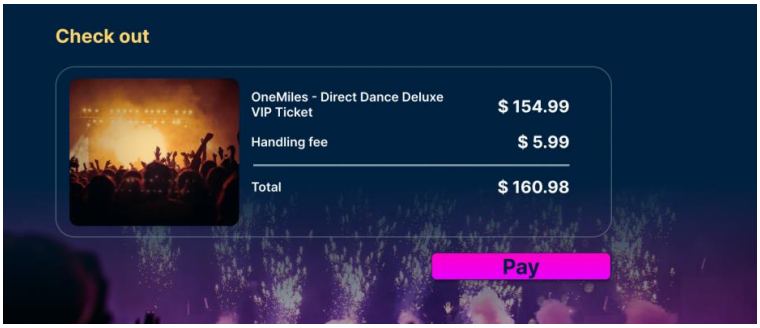




Hide (HD)

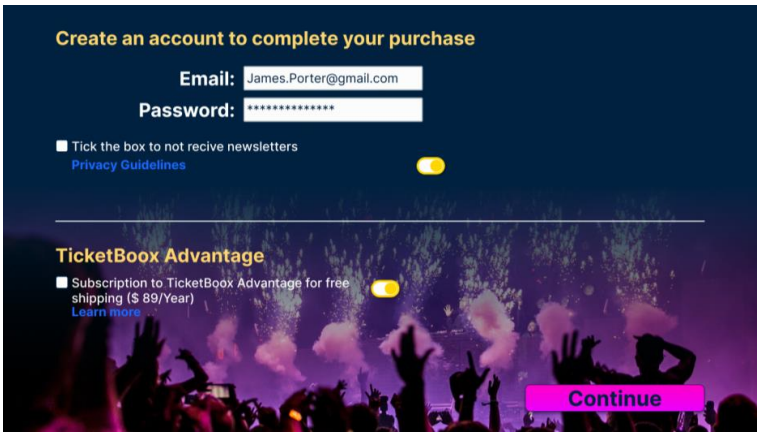

Get READY for the Beat.
OneMiles Performing Live in the City
June 15
Thursday, 18:30
Los Angeles, Colosseum
OneMiles - Direct Dance Deluxe VIP Ticket
$ 154.99
Buy Ticket
June 15
Thursday, 18:30
Los Angeles, Colosseum
OneMiles - Direct Dance Deluxe Standard Ticket
$ 74.99
Buy Ticket
June 15
Thursday, 18:30
Los Angeles, Colosseum
OneMiles - Direct Dance Deluxe Backseat Ticket
$ 55.99
Buy Ticket
A spectacle like this should not be missed. Rich beats enriched with pulsating melodies give the shows of OneMiles something of another world. Finely choreographed stage performances and brute effects amaze and fascinate critics and fans.

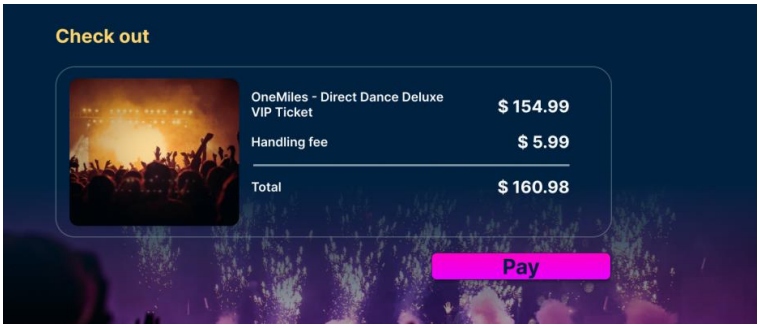

Create an account to complete your purchase
Email: James.Porter@gmail.com
Password: **********
Tick the box to not recive newsletters
Privacy Guidelines
TicketBoox Advantage
Subscription to TicketBoox Advantage for free shipping ($ 89/Year)
Learn more
Continue


Check out
OneMiles - Direct Dance Deluxe VIP Ticket
$ 154.99
Handling fee
$ 5.99
Total
$ 160.98
Pay

Switch (SW)


Get READY for the Beat.
OneMiles Performing Live in the City
June 15
Thursday, 18:30
Los Angeles, Colosseum
OneMiles - Direct Dance Deluxe VIP Ticket
$ 154.99
Buy Ticket
June 15
Thursday, 18:30
Los Angeles, Colosseum
OneMiles - Direct Dance Deluxe Standard Ticket
$ 74.99
Buy Ticket
June 15
Thursday, 18:30
Los Angeles, Colosseum
OneMiles - Direct Dance Deluxe Backseat Ticket
$ 55.99
Buy Ticket
A spectacle like this should not be missed. Rich beats enriched with pulsating melodies give the shows of OneMiles something of another world. Finely choreographed stage performances and brute effects amaze and fascinate critics and fans.


Create an account to complete your purchase
Email: James.Porter@gmail.com
Password: **********
Tick the box to not recive newsletters
Privacy Guidelines
TicketBoox Advantage
Subscription to TicketBoox Advantage for free shipping ($ 89/Year)
Learn more
Continue


Check out
OneMiles - Direct Dance Deluxe VIP Ticket
$ 154.99
Handling fee
$ 5.99
Total
$ 160.98
Pay

Bibliography

Luiz Adolpho Baroni, Alisson Andrey Puska, Luciana Cardoso de Castro Salgado, and Roberto Pereira. Dark Patterns: Towards a Socio-technical Approach. In *Proceedings of the XX Brazilian Symposium on Human Factors in Computing Systems*, IHC '21, pages 1–7, New York, NY, USA, October 2021. Association for Computing Machinery. ISBN 978-1-4503-8617-3. doi: 10.1145/3472301.3484336. URL https://doi.org/10.1145/3472301.3484336.

Aditi M. Bhoot, Mayuri A. Shinde, and Wricha P. Mishra. Towards the identification of dark patterns: An analysis based on end-user reactions. In *Proceedings of the 11th Indian Conference on Human-Computer Interaction*, IndiaHCI '20, page 24–33, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450389440. doi: 10.1145/3429290.3429293. URL https://doi.org/10.1145/3429290.3429293.

Kerstin Bongard-Blanchy, Arianna Rossi, Salvador Rivas, Sophie Doublet, Vincent Koenig, and Gabriele Lenzini. "I am Definitely Manipulated, Even When I am Aware of it. It's Ridiculous!" - Dark Patterns from the End-User Perspective. In *Designing Interactive Systems Conference 2021*, DIS '21, pages 763–776, New York, NY, USA, June 2021. Association for Computing Machinery. ISBN 978-1-4503-8476-6. doi: 10.1145/3461778.3462086. URL https://doi.org/10.1145/3461778.3462086.

John Brooke. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996. URL https://www.researchgate.net/profile/John-Brooke-6/publication/319394819_

SUS_--_a_quick_and_dirty_usability_scale/links/5ee5c59e458515814a5e6bb8/SUS--a-quick-and-dirty-usability-scale.pdf.

Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher. Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns. *Proceedings on Privacy Enhancing Technologies*, 2016(4):237–254, October 2016. ISSN 2299-0984. doi: 10.1515/popets-2016-0038. URL https://petsymposium.org/popets/2016/popets-2016-0038.php.

John P. Chin, Virginia A. Diehl, and Kent L. Norman. Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '88, pages 213–218, New York, NY, USA, May 1988. Association for Computing Machinery. ISBN 978-0-201-14237-2. doi: 10.1145/57167.57203. URL https://dl.acm.org/doi/10.1145/57167.57203.

Michael Chromik, Malin Eiband, Sarah Theres Völkel, and Daniel Buschek. Dark Patterns of Explainability, Transparency, and User Control for Intelligent Systems. *Los Angeles*, page 6, 2019. URL https://www.medien.ifi.lmu.de/pubdb/publications/pub/chromik2019iuiworkshop/chromik2019iuiworkshop.pdf.

Gregory Conti and Edward Sobiesk. Malicious interface design: exploiting the user. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 271–280, New York, NY, USA, April 2010. Association for Computing Machinery. ISBN 978-1-60558-799-8. doi: 10.1145/1772690.1772719. URL https://doi.org/10.1145/1772690.1772719.

Andrea Curley, Dympna O'Sullivan, Damian Gordon, Brendan Tierney, and Ioannis Stavrakakis. The design of a framework for the detection of web-based dark patterns. In *The Fifteenth International Conference on Digital Society*, ICDS 2021, pages 24–30, Nice, France, 2021. IARIA. URL https:

//arrow.tudublin.ie/cgi/viewcontent.cgi?article=1002&context=ascnetcon.

Linda Di Geronimo, Larissa Braz, Enrico Fregnan, Fabio Palomba, and Alberto Bacchelli. UI Dark Patterns and Where to Find Them: A Study on Mobile Applications and User Perception. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–14, New York, NY, USA, April 2020a. Association for Computing Machinery. ISBN 978-1-4503-6708-0. doi: 10.1145/3313831.3376600. URL https://doi.org/10.1145/3313831.3376600.

Linda Di Geronimo, Larissa Braz, Enrico Fregnan, Fabio Palomba, and Alberto Bacchelli. Ui dark patterns and where to find them: A study on mobile applications and user perception. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–14, New York, NY, USA, 2020b. Association for Computing Machinery. ISBN 9781450367080. doi: 10.1145/3313831.3376600. URL https://doi.org/10.1145/3313831.3376600.

Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. The Dark (Patterns) Side of UX Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 1–14, New York, NY, USA, April 2018. Association for Computing Machinery. ISBN 978-1-4503-5620-6. doi: 10.1145/3173574.3174108. URL https://doi.org/10.1145/3173574.3174108.

Colin M. Gray, Shruthi Sai Chivukula, and Ahreum Lee. What Kind of Work Do "Asshole Designers" Create? Describing Properties of Ethical Concern on Reddit. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, DIS '20, pages 61–73, New York, NY, USA, July 2020. Association for Computing Machinery. ISBN 978-1-4503-6974-9. doi: 10.1145/3357236.3395486. URL https://doi.org/10.1145/3357236.3395486.

Colin M. Gray, Jingle Chen, Shruthi Sai Chivukula, and Liyang Qu. End User Accounts of Dark Patterns as Felt Manipulation. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):372:1–372:25, October

2021. doi: 10.1145/3479516. URL https://doi.org/10.1145/3479516.

Paul Graßl, Hanna Schraffenberger, Frederik Zuiderveen Borgesius, and Moniek Buijzen. Dark and Bright Patterns in Cookie Consent Requests. *Journal of Digital Social Research*, 3(1):1–38, February 2021. ISSN 2003-1998. doi: 10.33621/jdsr.v3i1.54. URL https://jdsr.se/ojs/index.php/jdsr/article/view/54. Number: 1.

Johanna Gunawan, Amogh Pradeep, David Choffnes, Woodrow Hartzog, and Christo Wilson. A Comparative Study of Dark Patterns Across Web and Mobile Modalities. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):377:1–377:29, October 2021. doi: 10.1145/3479521. URL https://doi.org/10.1145/3479521.

Mairéad Hogan, Chris Barry, and Michael Lang. Dissecting Optional Micro-Decisions in Online Transactions: Perceptions, Deceptions and Errors. *ACM Transactions on Computer-Human Interaction*, May 2022. ISSN 1073-0516. doi: 10.1145/3531005. URL https://doi.org/10.1145/3531005.

Konrad Kollnig, Siddhartha Datta, and Max Van Kleek. I Want My App That Way: Reclaiming Sovereignty Over Personal Devices. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI EA '21, pages 1–8, New York, NY, USA, May 2021. Association for Computing Machinery. ISBN 978-1-4503-8095-9. doi: 10.1145/3411763.3451632. URL https://doi.org/10.1145/3411763.3451632.

Jamie Luguri and Lior Jacob Strahilevitz. Shining a Light on Dark Patterns. *Journal of Legal Analysis*, 13(1):43–109, January 2021. ISSN 2161-7201. doi: 10.1093/jla/laaa006. URL https://doi.org/10.1093/jla/laaa006.

Maximilian Maier and Rikard Harr. Dark design patterns: An end-user perspective. *Human Technology*, 16(2):170–199, August 2020. ISSN 17956889. URL https://www.proquest.com/docview/2681456628/abstract/5DAC6C1DF65345BBPQ/1. Num Pages: 170-199 Place: Jyväskylä, Poland Publisher: Centre of Sociological Research (NGO) Section: Articles.

Arunesh Mathur, Gunes Acar, Michael J. Friedman, Eli Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):81:1–81:32, November 2019. doi: 10.1145/3359183. URL https://doi.org/10.1145/3359183.

Arunesh Mathur, Mihir Kshirsagar, and Jonathan Mayer. What Makes a Dark Pattern... Dark? Design Attributes, Normative Considerations, and Measurement Methods. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, pages 1–18, New York, NY, USA, May 2021. Association for Computing Machinery. ISBN 978-1-4503-8096-6. doi: 10.1145/3411764.3445610. URL https://doi.org/10.1145/3411764.3445610.

Carol Moser, Sarita Y. Schoenebeck, and Paul Resnick. Impulse Buying: Design Practices and Consumer Needs. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–15, New York, NY, USA, May 2019. Association for Computing Machinery. ISBN 978-1-4503-5970-2. doi: 10.1145/3290605.3300472. URL https://doi.org/10.1145/3290605.3300472.

Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–13, New York, NY, USA, April 2020. Association for Computing Machinery. ISBN 978-1-4503-6708-0. doi: 10.1145/3313831.3376321. URL https://doi.org/10.1145/3313831.3376321.

Than Htut Soe, Cristiana Teixeira Santos, and Marija Slavkovik. Automated detection of dark patterns in cookie banners: how to do it poorly and why it is hard to do it any other way, April 2022. URL http://arxiv.org/abs/2204.11836.

Daniel Susser, Beate Roessler, and Helen Nissenbaum. Technology, Autonomy, and Manipulation. *Internet Pol-*

icy Review, 8(2):112–120, June 2019. URL https://papers.ssrn.com/abstract=3420747.

Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. (Un)informed Consent: Studying GDPR Consent Notices in the Field. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, CCS '19, pages 973–990, New York, NY, USA, November 2019. Association for Computing Machinery. ISBN 978-1-4503-6747-9. doi: 10.1145/3319535.3354212. URL https://doi.org/10.1145/3319535.3354212.

Christian Voigt, Stephan Schlögl, and Aleksander Groth. Dark patterns in online shopping: Of sneaky tricks, perceived annoyance and respective brand trust. In *HCI in Business, Government and Organizations: 8th International Conference, HCIBGO 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings*, pages 143–155. Springer, 2021. doi: https://doi.org/10.1007/978-3-030-77750-0_10. URL https://link.springer.com/chapter/10.1007/978-3-030-77750-0_10.

Index

Typeset August 22, 2023