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

RWTH AACHEN UNIVERSITY

Prevalent Dark Patterns: Investigating Children's Mental Models of Malicious Designs

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

by
Sarah Sahabi

Thesis advisor:
Prof. Dr. Jan Borchers

Second examiner:
Prof. Dr. Ulrik Schroeder

Registration date: 16.01.2023
Submission date: 07.07.2023

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

Our everyday interaction with the internet and mobile apps increasingly exposes us to deceptive practices and threats. These include dark patterns, which are malicious user interface elements that are designed to trick, confuse, or deceive users towards actions that may not be in their best interest. These practices pose risks to the users such as invasion of privacy and financial harm. While there has been extensive research on dark patterns in the context of adults, findings on children are still sparse. To develop effective countermeasures for this potentially vulnerable user group, it is crucial to gain insights into children's understanding of dark patterns. In our work, we investigated the following research questions: 1. What are children's mental models of dark patterns? 2. How is the prevalence of dark patterns reflected in these mental models? In our study at a local grammar school, 66 5th-graders aged 10-11 years completed a playful and multi-part questionnaire involving drawing, textual, and recognition tasks to elicit these mental models. Our findings suggest that children's mental models of dark patterns vary, with some demonstrating a general understanding of manipulative intents, while others have limited awareness. Children were more likely to recognise and use aesthetic manipulations compared to linguistic tricks or deceitful tactics. However, a clear relationship between prior exposure to dark patterns and mental models was not identified and requires further exploration. Our work lays the foundation for future research on the development of dark pattern countermeasures to ultimately safeguard children in the digital world.

Überblick

Durch unsere tägliche Interaktion mit dem Internet und mobilen Apps sind wir zunehmend betrügerischen Praktiken und Bedrohungen ausgesetzt. Dazu gehören sogenannte Dark Patterns, d. h. bösartige User Interface Elemente, die darauf abzielen, Nutzerinnen und Nutzer auszutricksen, zu verwirren oder zu Handlungen zu verleiten, die möglicherweise nicht in ihrem Interesse liegen. Diese Praktiken stellen Risiken, wie die Verletzung der Privatsphäre und finanzielle Verluste dar. Während das Gebiet im Zusammenhang mit Erwachsenen umfangreich erforscht ist, sind die Erkenntnisse über Kinder noch spärlich. Um wirksame Gegenmaßnahmen für diese potenziell gefährdete Nutzergruppe zu entwickeln, ist es entscheidend, Einblicke in das Verständnis von Kindern über Dark Patterns zu gewinnen. In unserer Arbeit haben wir die folgenden Forschungsfragen untersucht: 1. Was sind die mentalen Modelle von Kindern von Dark Patterns? 2. Wie spiegelt sich die Prävalenz von Dark Patterns in diesen mentalen Modellen wider? In unserer Studie an einem örtlichen Gymnasium füllten 66 Schülerinnen und Schüler der 5. Klassen im Alter von 10 bis 11 Jahren einen spielerischen und mehrteiligen Fragebogen mit Zeichen-, Text- und Wiedererkennungsaufgaben aus, um diese mentalen Modelle zu ermitteln. Unsere Ergebnisse deuten darauf hin, dass die mentalen Modelle der Kinder von Dark Patterns variieren, wobei einige ein grundlegendes Verständnis von manipulativen Absichten zeigen, während sich andere nur begrenzt darüber bewusst sind. Kinder erkannten und nutzten ästhetische Manipulationen häufiger als sprachliche Tricks oder betrügerische Taktiken. Ein eindeutiger Zusammenhang zwischen früherem Kontakt mit Dark Patterns und ihren mentalen Modellen konnte jedoch nicht festgestellt werden und bedarf weiterer Untersuchung. Unsere Arbeit bildet die Grundlage für künftige Forschungen zur Entwicklung von Gegenmaßnahmen für Dark Patterns, um schließlich Kinder in der digitalen Welt besser zu schützen.

Acknowledgements

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

Many thanks to René Schäfer and Annabell Brocker for your dedicated and motivating supervision of my work, as well as your valuable feedback, from which I have learned a lot.

Also a huge thank you to my family, friends, and all my i10 people. Thank you for always lending an ear, giving advice, and for your encouraging words.

Special thanks to Nicole and the Couven-Gymnasium Aachen for your immense support, collaboration, and for sharing my passion for the topic. Without your help, my work would not have been possible.

Conventions

Throughout this thesis we use the following conventions.

Text conventions

Definitions of technical terms or short excursus are set off in coloured boxes.

Excursus:
Excursus are detailed discussions of a particular point in a book, usually in an appendix, or digressions in a written text.

Definition: *Excursus*

The whole thesis is written in British English.

The first person is written in the plural form and unidentified third persons are referred to neutrally or in the plural form.

Chapter 1

Introduction

Since the dawn of the digital age, the rapid development of the internet within the last decades has changed the way we interact and access information [Lupiáñez-Villanueva et al., 2022]. As the primary interface to the digital world, mobile devices like smartphones and tablets have become indispensable in our everyday lives [Kollnig et al., 2021]. The mobility of devices gives constant access to the internet and enables global connectivity [Wellman et al., 2003], quick access to information [Leiner et al., 2009], and consumer welfare [Lupiáñez-Villanueva et al., 2022]. However, along with benefits, there has been an alarming rise in threats and deceptive practices on the internet [Kollnig et al., 2021, Lupiáñez-Villanueva et al., 2022]. Malicious trickery is used and UX best practices are deliberately inversed to manipulate and deceive users [Conti and Sobiesk, 2010]. One increasing form of online manipulation is the use of *dark patterns*.

The ubiquity of the internet comes along with malicious practices to deceive users

1.1 Dark Patterns

Definition: *Dark Patterns*

> **DARK PATTERNS:**
> Dark patterns are malicious user interface elements that are designed to trick, confuse, or deceive users towards actions that may not be in their best interest [Brignull et al., 2010, Gray et al., 2018].

Dark Patterns are researched in various scientific fields, and definitions and taxonomies vary accordingly

While today's definitions and classifications may differ by perspective and context, the term was introduced by UX researcher Harry Brignull in 2010 through his website devoted to exposing dark patterns and shaming companies that use them [Brignull et al., 2010]. Since then, researchers from various fields and backgrounds contributed to developing taxonomies and to classifying dark patterns under different aspects. This includes HCI [Mathur et al., 2021], ethics [Gray et al., 2018], law [Luguri and Strahilevitz, 2021], and behavioural economics [Lupiáñez-Villanueva et al., 2022]. The resulting taxonomies show that dark patterns encompass a wide spectrum of techniques. On one end, there are subtle emotional and linguistic tricks, such as misleading or emotionally manipulative formulations that influence users to choose options they may not have chosen otherwise [Lupiáñez-Villanueva et al., 2022]. On the other end, there are more aggressive or deceitful tactics, like the preselection of costly subscriptions that often remain unnoticed by the user [Mathur et al., 2019].

The prevalence of dark patterns poses privacy, financial, and mental threats to users

Falling for such dark patterns entails a number of risks for the user, including invasion of privacy [Di Geronimo et al., 2020], financial harm [Bongard-Blanchy et al., 2021], and mental harm [Lupiáñez-Villanueva et al., 2022]. The threat is exacerbated by the growing prevalence of dark patterns in UIs: today, they cover a wide range of popular websites and apps, from shopping and booking sites [Mathur et al., 2019] to social media platforms [Mildner and Savino, 2021]. They have even become so ingrained in our everyday lives that users no longer notice them [Di Geronimo et al., 2020, Conti and Sobiesk, 2010]. However, even when users are aware of dark patterns, in many cases, they are unable to resist them [Bongard-Blanchy et al., 2021].

With the rise of smartphones, a whole range of new dark patterns specifically tailored for mobile devices emerged [van Nimwegen and de Wit, 2022]. Today, dark patterns are even more prevalent in mobile contexts compared to other platforms [Gunawan et al., 2021]. Given the pervasive role of mobile devices in everyday life [Kollnig et al., 2021], it is essential to examine the presence of dark patterns in mobile apps. This is especially important considering that approximately 95% of the most popular apps employ such deceptive techniques [Di Geronimo et al., 2020]. Consequently, users may encounter them particularly frequently while being around twice as susceptible to them in the mobile context as compared to the desktop context [van Nimwegen and de Wit, 2022]. Apps that are specifically often affected by dark patterns are mobile games [Bell and Fitton, 2021]. However, due to the vast number of dark patterns they incorporate, they pose a major threat to their users, among them vulnerable groups like children.

Dark patterns are particularly prevalent in mobile contexts, e.g. mobile games

1.2 Children's Vulnerability to Dark Patterns

Several factors suggest that children could be particularly vulnerable to dark patterns and the associated risks: first, research has shown that children are in general more easily manipulable than adults [Valkenburg and Piotrowski, 2017]. Second, recent statistics about German children [Rohleder, 2022] show that they often lack experience in using the internet and dealing with manipulative practices both online and offline. At the same time, they gradually move towards less and less supervised use. Finally, because children are often driven by curiosity and exploration, they are rarely inhibited by fears such as privacy concerns; they think of mistakes as part of the learning experience rather than being cautious to avoid them [Bell and Fitton, 2021, Morrison et al., 2021]. The increased susceptibility of mobile users to dark patterns [van Nimwegen and de Wit, 2022] puts children in an even more vulnerable position, as they primarily access the internet using mobile devices [Rohleder, 2022]. Also, statistics indicate that more than

Children are vulnerable to dark patterns as they are more easily manipulable and more inexperienced and daring in internet use

90% of German 10–12-year-olds use mobile devices regularly [Rohleder, 2022]. Taking into consideration that they spend an average of 1.5 hours a day online — often unsupervised [Rohleder, 2022] — the topic of children and dark patterns in the sense of safe media use is becoming urgently relevant.

Children as a target group of dark patterns are relatively unexplored

Nevertheless, while there is already extensive research on how adults perceive and interact with dark patterns (e.g., Di Geronimo et al. [2020], Lupiáñez-Villanueva et al. [2022]), the field of children's understandings of online manipulation, privacy, and security is still relatively unexplored [Morrison et al., 2021, Kumar et al., 2017, Brodsky et al., 2021]. Our aim for this work is to address this research gap by exploring children's mental models of dark patterns, i.e. gaining a better understanding of whether and how children make sense of dark patterns when they encounter them. These findings could help to design effective and much-needed interventions that meet children's needs, which may differ from those of adults [Bruckman et al., 2012, Druin, 1999]. As today's children have grown up with smart mobile devices, we also want to explore the extent to which the prevalence of dark patterns in popular apps is reflected in their mental models. This leads us to the following research questions:

We take a first step towards understanding children's mental models of dark patterns

RQ1: What are children's mental models of dark patterns?

RQ2: How is the prevalence of dark patterns reflected in these mental models?

We aim to answer these two questions with a mental models research approach.

1.3 Mental Models

Mental models research is an interdisciplinary field

Mental models research centres around the elicitation of mental models of a target group for a system or concept of interest. It is used in many different disciplines, mainly in HCI [Norman, 1983], cognitive science [Johnson-Laird, 1980], and system dynamics [Doyle and Ford, 1998]. Despite its prominence among researchers, there is no consistent, explicit definition of what a mental model is [Doyle and Ford, 1998]. In the course of this work, we use the following definition of mental models, derived from Norman [1983], Staggers and Norcio [1993], and Norman [2002]:

MENTAL MODELS:
Mental models are abstract models in people's minds of what they believe something is, how it works, and how to interact with it.

Definition:
Mental Models

In the context of this work, mental models of dark patterns involve children's perception and understanding of deceitful design choices on a website or in an app, and their awareness of the intent behind them.

1.4 Outline

This thesis aims at understanding children's mental models of dark patterns using a specially created multi-part questionnaire which was conducted in multiple 5th-grade classes of a local grammar school.

We first provide a background of existing work on dark patterns in general, as well as children's understanding of internet security, in Chapter 2 "Related Work". On the basis of insights from mental models research and studies with children, we proceed in Chapter 3 "Mental Models Study" with an extensive discussion of the creation and considerations of our mental models questionnaire, as well as the methods used when conducting the study and analysing the data. The results for all parts of the questionnaire are then individually presented, with a focus on different viewpoints to understanding mental models.

In Chapter 4 "Discussion", we discuss our findings and their implications, compare them to previous work, and point out limitations of the questionnaire and study procedure. Finally, we conclude with a summary of the contributions of this thesis and propose open questions for future work building up on our findings in Chapter 5 “Summary and Future Work”.

Chapter 2

Related Work

In the following sections, we provide an overview and discuss related work in the field of dark patterns. We also present what has been researched in the context of children and their understanding of internet security, and which questions remain open.

2.1 Dark Patterns

We start by reviewing the literature on existing dark pattern taxonomies and their occurrences in our everyday lives. We then proceed with discussing their impact on users and possible approaches to countering them.

2.1.1 Taxonomies

As the prevalence of dark patterns increases, it becomes more and more vital to gain a better understanding of them. Categorising them and creating taxonomies facilitates comprehension and helps to raise awareness and identify open questions, such as how to combat particular dark patterns [Lupiáñez-Villanueva et al., 2022]. Taxonomies can promote research in the field and allow for exchange and communication through well-defined and consistent terminol-

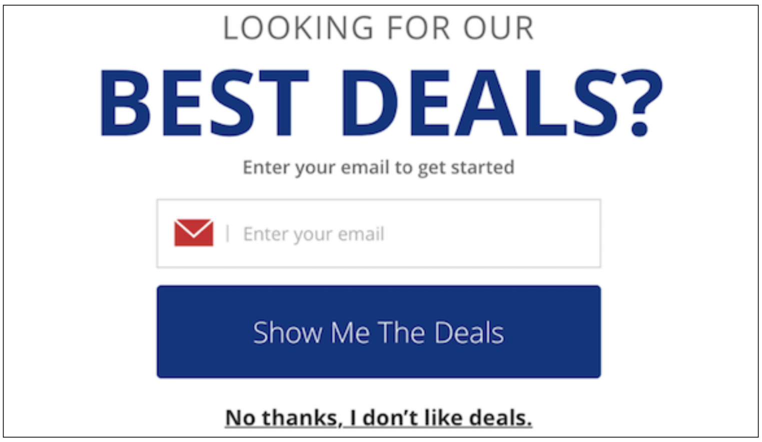

Figure 2.1: An example for the Confirmshaming dark pattern, taken from `www.sears.com/`: users might be tempted to enter their emails because of the guilt-inducing wording of the reject option.

Different taxonomies exist depending on background and context

ogy [Mathur et al., 2019]. However, just as there is no uniform definition of the umbrella term *dark patterns* [Gray et al., 2018], there is also not one universal dark pattern taxonomy [Lupiáñez-Villanueva et al., 2022]. Existing taxonomies differ in their viewpoints and areas of application, as well as in the data and definitions they have been built on [Mathur et al., 2019, 2021]. A collection of dark pattern contained in those taxonomies can be taken from Appendix A "A Collection of Dark Pattern Types".

The first taxonomy was created by Brignull et al. [2010] and contained 12 patterns

Brignull et al. [2010] took a first step towards the creation of such taxonomies by collecting and analysing incidents of deceptive designs on the internet. They categorised them into recurring patterns and published these under the name of *dark patterns* on their website. For this, they identified a total of 12 dark pattern types, among them *Confirmshaming*, where targeted strategies are used to induce shame or guilt, thus emotionally manipulating users into making certain decisions that they might not have made on their own (see example in Fig. 2.1). Another type introduced by Brignull et al. [2010] is *Trick Question* (recently renamed to the more general term *Trick Wording*), which misleads users to take an action by using confusing language, e.g., ambiguous wording or double-negatives.

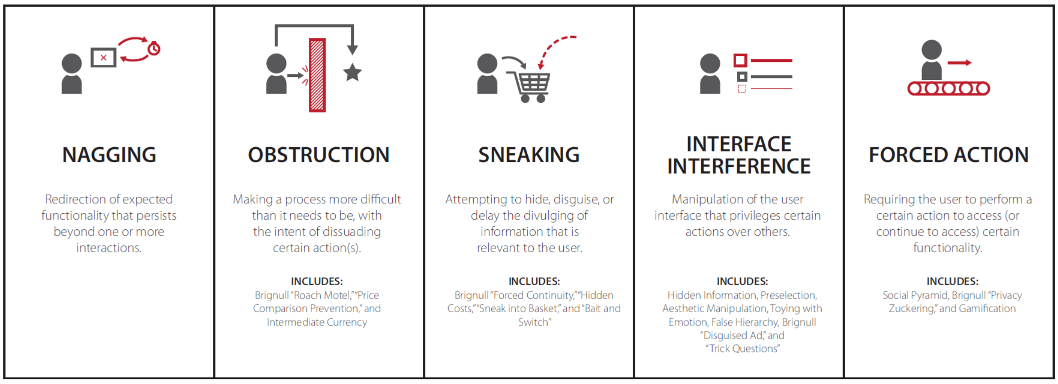

Figure 2.2: Gray et al. [2018] extended previous taxonomies and identified the five superordinate categories *Nagging, Obstruction, Sneaking, Interface Interference,* and *Forced Action*. Pattern names within quotation marks originate from Brignull et al. [2010]'s initial taxonomy. Figure taken from Gray et al. [2018].

At around the same time, Conti and Sobiesk [2010] proposed a taxonomy of malicious interface design techniques based on a 12-month analysis of websites, applications, and interfaces off the desktop. To validate their initial findings, they conducted a study in which participants pointed out all malicious designs they could find in a given set of on and off-desktop contexts. Finally, they completed their list by actively searching for previously unidentified techniques through a group discussion with hacking-experienced participants. Although at that time, the term *dark pattern* was not yet introduced and different data gathering and analysis strategies were used, the two taxonomies of Brignull et al. [2010] and Conti and Sobiesk [2010] show partial overlaps. For example, the malicious technique *Confusion* complies with Brignull et al. [2010]'s definition of the previously mentioned *Trick Wording*.

The taxonomy of Conti and Sobiesk [2010] was developed in parallel to Brignull et al. [2010]'s, even before the word *dark pattern* was introduced

While the dark pattern types from the taxonomies presented are still current and widely used, more and more new manipulative designs have emerged in recent years, necessitating continuous updates of the categories [Narayanan et al., 2020, Lupiáñez-Villanueva et al., 2022]. A widely recognised and more contemporary taxonomy was composed by Gray et al. [2018] who extended Brignull et al. [2010]'s work by new pattern types and rearranged them into five superordinate categories: Nagging, Obstruc-

Gray et al. [2018] extended existing taxonomies by the five superordinate categories: Nagging, Obstruction, Sneaking, Interface Interference, and Forced Action

tion, Sneaking, Interface Interference, and Forced Action (see Fig. 2.2). For this, they compiled a corpus of dark pattern occurrences from a large number of popular online platforms. Then, they structured the data in terms of the designers' underlying motivations, both through an open coding approach and through constant comparative analysis of existing types. For example, while Brignull et al. [2010]'s *Confirmshaming* aims at triggering emotions of shame, often implemented through dedicated formulations, Gray et al. [2018]'s *Toying with Emotions* is an extended version of this type of pattern. It denotes the use of language, style, colour or the like to trigger any kind of emotion intended to influence the user into an action. This pattern was classified as *Interface Interference*, together with the newly defined *Preselection*, where users face already preselected default options, e.g., ticked checkboxes. This pattern exploits the chance of users overlooking this and, thus, tricks them, for example, into accepting something they did not want.

Mathur et al. [2019] developed a taxonomy specifically for shopping websites

Over time, the taxonomy landscape has also been expanded to include dark pattern taxonomies for specific contexts and applications [Mathur et al., 2021]. For instance, Zagal et al. [2013] focused on dark patterns in video games, while Mathur et al. [2019] took a closer look specifically at shopping websites and grouped the patterns according to a set of dark pattern characteristics. With this, they aimed at depicting the influence on users' decision-making and potential harm per dark pattern. While their findings substantially built on earlier contributions, their work mainly differed from most prior work in its data collection method: instead of basing their research on anecdotal data or user submissions, Mathur et al. [2019] relied on a web crawler – an automated technique to identify instances of dark patterns from $\sim$11K shopping websites. Their work resulted in several new types, including *Visual Interference*, a pattern that is described as using visual elements to interfere with the page design to influence users into making certain decisions, e.g., being tempted to click on the large, colourful button instead of the small, low-contrast alternative.

2.1.2 Prevalence of Dark Patterns

Having established a repertoire of dark pattern taxonomies and types, various researchers proceeded to assess the occurrence of such patterns in the wild. Using their web crawler, Mathur et al. [2019] were not only able to show the general prevalence of dark patterns on shopping websites, but they also found that more popular websites were more likely to contain dark patterns.

Researchers like Mathur et al. [2019] and Di Geronimo et al. [2020] showed the prevalence of dark patterns in apps and websites

Di Geronimo et al. [2020], on the other hand, focused their research on the prevalence of dark patterns specifically in mobile apps. From eight main app categories of the Google Play Store, they selected the 30 most trending apps per category and collected all instances of dark patterns they could find for each in a publicly available dataset[1]. In contrast to previous work with similar research questions, they examined the apps through active interactions while performing a series of tasks rather than basing their results on screenshot interactions. Most notably, they found that among the 240 apps, 95% contained at least one dark pattern and that the average number of occurrences per app was more than seven. Moreover, they discovered differences between different app categories regarding the number of dark pattern occurrences. For instance, they reported that the 'News and Magazine' category incorporated fewer dark patterns than other types, such as 'Entertainment', 'Shopping', and 'Communication'. Overall, the three most prevalent dark pattern types were Nagging, False Hierarchy, and Preselection.

Dark patterns were prevalent in 95% of the most prominent apps tested by Di Geronimo et al. [2020] and vary in frequency depending on the app category

Lupiáñez-Villanueva et al. [2022] explored the same question of dark pattern occurrences in different websites and apps and delved even further into the topic by comparing the intensities of prevalence across EU countries. In a study with participants from different EU countries, all manipulative commercial practices found on the most popular websites and apps in the EU were identified and documented. Similar to Di Geronimo et al. [2020], they determined that 97% of those interfaces contained at least one dark pattern

Dark patterns were roughly equally prevalent among prominent apps and websites across different EU countries

[1] https://figshare.com/s/048c984854a59429d0f0
Accessed: May 10, 2023

and that most interfaces comprised combinations of several dark patterns at once. Besides this, the researchers did not detect clear differences in the prevalence levels across different countries in and outside of the EU, nor between mobile apps and websites.

Gunawan et al. [2021] determined that most dark patterns were more prevalent in mobile apps than in mobile or web browsers

The latter, however, conflicts with subsequent findings by Gunawan et al. [2021]. While previous work only loosely considered the effect of different modalities, Gunawan et al. [2021] conducted a comparative study of dark pattern prevalence levels of 105 popular services across the three modalities mobile application, mobile browser, and web browser. Contrasting prior work, they indeed detected a difference in the prevalence between modalities, as 30 out of 46 dark patterns occurred more frequently in apps than in mobile or web browsers. Moreover, they state that they had found discrepancies between their own dark pattern frequencies and those reported by Di Geronimo et al. [2020], which, as they conclude, underlines the challenges in consistent dark pattern measurement based on different strategies [Gunawan et al., 2021].

2.1.3 Impact on the User & Awareness

"Any short-term gains a company gets from a dark pattern is lost in the long term."

— Hoa Loranger,
Vice President Nielsen Norman Group[2]

Cotte et al. [2005] detected negative feelings of users towards brands when encountering malicious designs

Already long before the term *dark pattern* was prominent, Cotte et al. [2005] discovered in a study that consumers harboured negative feelings towards advertising brands as soon as they perceived manipulative intentions. Furthermore, Conti and Sobiesk [2010] investigated users' self-reported frustration and tolerance of common malicious interface designs depending on the context and the task they were trying to accomplish. The results suggest that participants found all malicious designs significantly frustrat-

[2] `https://www.fastcompany.com/3060553/why-dark-patterns-wont-go-away` *Accessed: May 12, 2023*

ing. However, for different tasks or activities, they demonstrated varying levels of tolerance. For example, higher tolerance for frustration was shown in gaming, shopping, and pornographic applications, and the lowest tolerance for search, news, weather, and vendor support sites.

Another dimension of attitude towards such designs was later studied by Luguri and Strahilevitz [2021] who were interested in user annoyance with dark patterns. They distinguished between two groups of dark patterns: mild patterns and aggressive patterns. In a controlled experiment, dark patterns of both groups were embedded in a fictitious website attempting to persuade users to purchase an insurance policy against identity theft. Participant behaviour and responses were then collected while interacting with the website. Aggressive dark patterns were shown to trigger a fierce backlash and annoyance. Mild patterns, in contrast, did not. The researchers also identified differences between the two groups with regard to *preference inconsistency*, i.e., the effectiveness of the manipulation to beguile a person into making choices they would not have made themselves. While mild patterns were two times more effective than the usual user interface from the control conditions, aggressive patterns were even four times more effective.

Luguri and Strahilevitz [2021] found that aggressive dark patterns received more backlash and were more effective than mild dark patterns

Lupiáñez-Villanueva et al. [2022] expanded on the comprehension of preference inconsistency. In an online study, they tested the impacts of dark patterns on the decision-making of people across the EU. About 7500 participants from different EU countries were asked to choose from two different digital entertainment service packages from a web page that contained various dark patterns. Then, the consistency between their selection and their previously stated preference was measured. To gain more insights, time pressure was included for one group of participants as another factor. The experiment revealed a set of pattern types that triggered higher degrees of preference inconsistency, among them *Hidden Information* and *Toying with Emotions*. What is more, they found that not all participants were equally vulnerable to dark patterns. For example, participants from the group that was exposed to time pressure showed stronger inconsistencies. Similar observations

Lupiáñez-Villanueva et al. [2022] observed differing degrees of preference inconsistencies for different types of patterns and different participant groups were made with older or less educated participants.

Di Geronimo et al. [2020] revealed that users performed better at detecting dark patterns when they were informed about them before

Finally, researchers investigated the relationship between awareness of dark patterns and users' capabilities to detect them in a design. In Di Geronimo et al. [2020]'s online experiment, participants were asked to describe any malicious designs they could detect in a short video of an interaction with an app. The majority of participants stated that they either could not detect any, or they were not sure about their answers. One of the reasons that emerged from subsequent exploratory investigations was that the prevalence and commonality of dark patterns made them hard to detect as they had become part of normal everyday interactions. Accordingly, Lupiáñez-Villanueva et al. [2022] realised that users accepted and had become accustomed to their presence so they would no longer notice them. Nevertheless, once participants became aware of the use of dark patterns and became more knowledgeable about them, they performed better at detecting dark patterns [Di Geronimo et al., 2020].

Studies on the effect of awareness on preference inconsistency yield contradictory results

Additionally, a series of projects were dedicated to research on how preference inconsistency was affected by users' awareness of dark patterns. However, findings from different works diverged partly. For instance, Grossklags and Acquisti [2007] discovered that participants were more likely to resist the persuasive power of malicious designs to some extent when they were aware of their presence. Bongard-Blanchy et al. [2021], on the other hand, noted that although some participants of their experiment were aware of the presence of dark patterns, this was not reflected in their behaviour or ability to resist the manipulation. The authors suspected this was because users might not have been aware of the actual harm and dangers that emanated from the manipulations. Conclusively, although not always reflected in the users' decisions, researchers agree that being educated and aware of such prevalent dark patterns might be a powerful approach to becoming more sensitive and, in the long run, more protected against such prevalent dark patterns [Brignull et al., 2010].

2.1.4 Fighting Dark Patterns

Raising awareness is an effective countermeasure against falling for dark patterns

Given the prevalence and the strong influence of dark patterns on people, researchers emphasise the urgent need for effective countermeasures [Conti and Sobiesk, 2010, Bongard-Blanchy et al., 2021]. As discussed earlier, creating awareness among people is a promising approach to combat. This can apply to the end user, e.g. through dedicated public denouncing websites like Brignull et al. [2010]'s, as well as to the responsible advertisers and designers, e.g., through ethics education [Gray et al., 2018].

Tools to manually counteract dark patterns require expertise and awareness

Beyond that, there are a number of other strategies to protect against dark patterns, the most predominant of which will be reviewed in the following. With the onset of the relevance of the topic, Conti and Sobiesk [2010] studied the effectiveness and ease of use of existing countermeasures taken by expert users at the time. These included the use of personal proxies, pop-up blockers, text-only browsers, adblocking software, and browser plug-ins. However, their analysis showed that existing methods were not adoptable by non-expert users and their effectiveness was limited. A more recent approach is presented by Kollnig et al. [2021] whose community-driven app modification framework allows users to manually disable dark patterns in apps in a controlled way. While this comes with the advantage that users need no particular expertise to use it, it requires a certain level of awareness to recognise dark patterns in the first place — which is seldom given [Di Geronimo et al., 2020]. Researchers, therefore, concluded that more focus should be put on developing effective and easy-to-use tools that can reliably detect dark patterns and protect the user from their impact [Bongard-Blanchy et al., 2021, Conti and Sobiesk, 2010].

In fact, there have been a variety of projects dedicated to the automatic detection of dark patterns. Within the *Dark Pattern Detection Project*[3], researchers attempted to achieve this through AI-based text analysis methods. The resulting 'Dark Pattern Detection App' recognises text-based dark patterns and redesigns the page in a way to highlight the

[3] `https://dapde.de` *Accessed: May 10, 2023*

AI-based automated dark pattern detection tools are currently limited in the dark pattern types they can detect

manipulation to warn the user early on. An early prototype of this tool was introduced by Hausner and Gertz [2021] through an example within the domain of cookie banners. However, the countermeasure is restricted to text-based manipulations only and hence does not include a variety of other dark pattern types.

Curley et al. [2021] classified dark patterns according to how well they could be detected automatically and propose a framework for detection tools which considers these different classes

To identify the challenges of automatic detection through machine learning, Soe et al. [2022] trained a supervised machine learning model on a manually collected dataset of 300 cookie banners. Subsequent tests revealed moderate detection accuracies and were, again, limited to specific domains and pattern types only. Further review of other work on automatic dark pattern recognition also yields similar limitations: there is no known system that automatically detects every type of dark pattern. Curley et al. [2021] took a step back and looked from a theoretical point of view into the question of whether it is technically possible to automatically detect all types of dark patterns. As they found that some pattern types could be recognised more reliably than others, and others were impossible to recognise, they propose a framework for dark pattern detection consisting of the following three classes: 1) patterns that can be detected in an automated way (e.g. Trick Question), 2) patterns that can be detected in a manual way (e.g. Hidden Cost), and 3) patterns that cannot be detected (e.g. Confirmshaming). As the authors suggest, this classification can be used in the development of an appropriate detection tool, where class 1 patterns are detected and removed and potential class 2 patterns are highlighted to warn the user.

Several directives against the usage of dark patterns have already been legislated and even more are requested

A completely different perspective on the issue of dark pattern countermeasures is the law. At present, there exists already a series of laws to protect humans from the manipulative power of dark patterns [Berbece, 2019]. EU legislation to regulate dark patterns includes, among others, the Unfair Commercial Practices Directive (UCPD), Consumer Rights Directive (CRD), Unfair Contract Terms Directive (UCTD), and the General Data Protection Regulation (GDPR) [Lupiáñez-Villanueva et al., 2022]. A collection of existing laws related to each dark pattern was compiled and published by Brignull et al. [2010] on their website. For instance, the use of the dark patterns *Hidden Cost* and *Sneak into Basket* has become illegal in several countries of the EU through the CRD. However, despite existing laws, more and more experts in the field are demanding further protection laws [Weinzierl, 2020, Jarovsky, 2022].

Overall, a number of researchers have devoted themselves in recent years to combating the dangers posed by dark patterns, using a wide variety of approaches. Nevertheless, the challenge remains that new patterns will continue to emerge, which places the demand on systems to always be reactive and dependent on constant updates of existing dark pattern databases [Hausner and Gertz, 2021].

2.2 Children's Mental Models and Awareness of Digital Security

Having looked at related literature focusing on dark patterns, this section will cover work on children and their understanding of digital security. In particular, we will review work on children's mental models and awareness of security and privacy both in the context of the internet and mobile apps. We will also explore literature in the field of children and dark patterns, especially in the context of mobile games.

Mental models research as an approach to identify risks of children in digital security

As explained earlier, with the rising digital activity of children [Rohleder, 2022], the research on children's digital security becomes more and more urgent [Morrison et al., 2021]. Mental models research can be a powerful approach for revealing the weak spots of children regarding digital security [Brodsky et al., 2021]. It can also disclose novel insights that help establish how to counteract these threats [Papastergiou, 2005].

Just as Zhang [2008] and Dinet and Kitajima [2011] did before, Brodsky et al. [2021] investigated children's mental models of the internet to gain a first impression of children-specific risks. From these models, the latter group of researchers discerned that children were mostly aware of the ubiquity of the internet. However, they primarily thought

Children's mental models of the internet hinted at online privacy and security risks

of traditional devices with which they could directly interact with the internet and only rarely considered Internet of Things devices. The authors pointed out that this might indicate a lack of understanding of how data could be collected and used by everyday objects. Therefore, they concluded that the dearth of Internet of Things devices in children's mental models suggested serious online privacy and security risks for them. Consequently, they argued that it was crucial for children to develop awareness and understanding of this to protect themselves against such risks.

Mental models can help develop appropriate educational programs

Papastergiou [2005] explored this particular problem and used mental models as a starting point to construct measures for enhanced internet security of children. From the mental models of the internet which they extracted from 340 high school children through questionnaires and drawings, they ultimately derived implications for the tailored development of didactical methods in informatics education.

Delving deeper into the topic of children's mental models of privacy and security, Zhang-Kennedy et al. [2016], as well as Oates et al. [2018], found that when asking children to draw what privacy meant to them, they mainly thought of physical privacy, e.g. in bathrooms, bedrooms, etc. Only a few children mentioned digital privacy. In contrast, when repeating the same study with adults, both research groups made the same observation that adults were thinking more about digital privacy and internet security. For example, they mentioned online predators, cyberbullies, and cybercriminals [Zhang-Kennedy et al., 2016].

Children are aware of some online privacy and security risks but need to be better educated by their parents about how to protect themselves

However, Kumar et al. [2017] accomplished to show that when children were explicitly asked to express their mental models of *online* privacy and security, they indeed proved to have a general understanding of it. Zhao et al. [2019] expanded on this, as they reported that children were aware of some online privacy risks, such as oversharing information, and could articulate them well. Other risks, however, they were not aware of, which again highlights the necessity of education. Moreover, Kumar et al. [2017] discovered that children already implemented some strategies to protect themselves online, e.g. by entering wrong answers when asked about personal information. Notably, they found that children strongly relied on their parents' advice when they were unsure about whether something was safe or not. With this, the authors emphasise the important and powerful responsibility of parents to educate their children. However, as the authors pointed out, there is also a need for guidance for parents to teach children the correct and safe online behaviour, for instance through educational programs developed by experts.

In the context of children's digital security, some researchers see particularly high risks in mobile games. These often contain dark patterns to increase monetisation and are especially popular with children [Bell and Fitton, 2021]. To inform discussion and provide starting points to develop effective countermeasures, Fitton and Read [2019] took a closer look at the incidence of dark patterns in free-to-play apps. Based on data from a literature review and a qualitative study with 39 school children, they composed a framework of dark pattern categories that children frequently encountered in mobile games. These include the categories of monetary, social, and inappropriate dark patterns. To expand the knowledge and identify more concrete threats to children from dark patterns in mobile games, Bell and Fitton [2021] undertook a literature survey. Building on previous findings about children's limited understanding of risks in digital environments, they suspected negative impacts of high exposure of children to dark patterns. Specifically, they identified risks related to commerce, (harmful) content, contact and conduct, i.e. becoming perpetrators themselves. They concluded that there has been too little focus on creating awareness among children and that this, along with necessary regulations, would be an important next step in protecting children from dark patterns.

Dark patterns in mobile games pose a particular threat to children, related to commerce, content, contact, and conduct

As mentioned earlier, when developing new educational programmes, it is important to first understand the existing mental models of the target group [Papastergiou, 2005]. With this aim in mind, Morrison et al. [2021] proposed the design of an ethical informed workshop for revealing children's mental models of online dark patterns. Their method comprised drawing tasks where children would be given screenshots of a scenario which contains a dark pattern. The task was, to draw what they thought would hap-

Understanding children's mental models of dark patterns might provide new ways to protect them pen if someone would click on a specific button or link that was part of the dark pattern. They extended the methodology by a discussion of further questions about the drawings and possible risks the children might anticipate from engaging by clicking. This would enable the researchers to understand the reasoning behind the drawings and gain deeper insights. In follow-up work, partly by the same authors, Clift et al. [2022] tested this methodology in a remote setting (due to the COVID-19 pandemic) in the classroom, and concluded the validity of their method.

Up to now, there is no work about how children identify and understand manipulative design elements

In the described methodology, however, dark patterns are considered in isolation from the remaining UI and the children are explicitly pointed to dark pattern elements and asked about their impact. Although we believe this may provide important insights into whether children can understand the dangers of dark patterns, we think it is necessary to also consider mental models of dark patterns as part of a whole interface and explore whether children can identify manipulative design elements on their own. Hence, this requires a method that does not direct and limit questions to specific elements of the UI. Since, to the best of our knowledge, there is no work that has investigated this, we will address this research gap within this thesis.

Chapter 3

Mental Models Study

This chapter contains a detailed description of our methodology for the mental models study with children. We will then conclude with a presentation of our results.

3.1 Method

The Methodology section includes a detailed overview of the questionnaire design and the procedure of the study with school classes. We will proceed with the demographics of our child participants and discuss the resulting aggravated ethical considerations we needed to comply with. Finally, we will discuss the data analysis methods we used for the qualitative and quantitative data.

3.1.1 Study Design Considerations

Choosing the right study design for mental models elicitation can be challenging. Even more so, when it involves children as the target group. In the following two sections, we review relevant literature and highlight special considerations that need to be taken into account in studies with children as well as in mental model elicitation.

Studies with Children

The design of a study with children requires to be especially carefully considered by the research team [Bruckman et al., 2012]. The most important aspects to consider will be presented in the following.

Studies with children usually involve ethical challenges

First, one of the most discussed issues is the ethical restrictions that arise in studies with minors [Punch, 2002]. This is one of the reasons why children are considerably unexplored [Morrison et al., 2021]. Particularly sensitive aspects of this are issues of informed consent, confidentiality, and the child's comfort, which must be guaranteed at all times [Punch, 2002]. Corresponding ethical guidelines will be discussed in more detail in Section 3.1.5 "Ethical Considerations".

Child participants have different needs compared to adults and the study must be designed accodingly

Second, it is crucial to be aware of the differences between children and adults in research, as well as the challenges those imply [Bruckman et al., 2012]. Above all, it must be taken into account that children are not 'miniature adults' and have their own needs and expectations with regard to the study design [Marhan et al., 2012, Punch, 2002]. A major difference between children and adults is the generally more limited understanding and vocabulary of children [Boyden and Ennew, 1997, Punch, 2002]. Therefore, it is essential to use age-appropriate language only [Morrison et al., 2021]. Furthermore, it is critical to choose appropriate research methods that balance the lower attention span and potential nervousness of the child participant, e.g. fun and child-friendly methods [Boyden and Ennew, 1997, Punch, 2002]. This is especially important, as Prokop et al. [2008] state that the type of instruction strongly affects the child's responses.

Children are usually more prone to demand characteristics

Another aspect to consider is that children are often more affected by demand characteristics, which poses a threat to the validity and reliability of the study [Punch, 2002, Norman, 1983]. Researchers pointed out that children may be afraid of adults' reactions to what they say because they are used to trying to satisfy them, which might eventually lead them to lie [Punch, 2002]. For example, conducting the study in a school environment might cause children to feel pressured to give the correct answer to all questions because that is what they are usually expected to do in school. To counteract this, researchers must emphasise throughout the study that there are no right or wrong answers.

Finally, the analysis of children's responses often poses a methodological problem [Punch, 2002]. Adult researchers tend to interpret children's statements and make their own potentially incorrect assumptions in the belief that they know and understand them [Punch, 2002, Fine and Sandstrom, 1988]. The only way to counteract this is to always try to remain objective and uninfluenced in the analysis.

Researcher bias occurs often in research with children

Existing Mental Model Elicitation Techniques

To gain deeper insights into children's understanding of dark patterns, we decided to take a mental models approach for this thesis. As defined in Section 1.3 "Mental Models", mental models describe users' conceptions of what something is and how it works. They are formed through interactions and have predictive and explanatory power, guiding our decisions on how to interact with the world [Norman, 1983, Staggers and Norcio, 1993, Norman, 2002]. Understanding people's mental models gives researchers valuable insights into how systems should be implemented to be compatible with existing, potentially flawed user understanding and to enable a positive, error-free user experience [Norman, 1983, Kodama et al., 2017].

Understanding mental models is crucial to optimised user experience of a system

However, the elicitation of mental models can be challenging, as they are often incomplete, unscientific, and indistinct [Norman, 2002, Kodama et al., 2017]. While these models likely differ between people for the same construct, even a single person might have multiple distinct models of the same construct [Norman, 1983]. The research field around mental models offers a variety of different elicitation techniques whose use cases have been widely studied and discussed. These include interviews, questionnaires, and drawings [Marhan et al., 2012]. Depending on the context of the research and the target population, different

Mental model elicitation is challenging and requires appropriate methods depending on the context techniques may be more or less suitable. As explained in the previous section, it is essential to choose the methodology carefully when conducting research with children. In the following, we provide an overview of existing elicitation techniques and discuss their advantages and disadvantages, particularly with regard to our research question about children's mental models of dark patterns.

Interviews are comparatively easy to evaluate, but questionable in terms of the validity of the results.

Interviews: The most commonly used technique for eliciting mental models is interviewing, which belongs to the category of direct elicitation techniques [Doyle et al., 2022, Marhan et al., 2012, Jones et al., 2011]. This means that models are defined by the participants themselves and do not rely on the researchers' interpretations. It is therefore comparatively uncomplicated to implement and analyse. At the same time, this is one of the most controversial techniques, especially when applied to children; the lack of self-awareness [Marhan et al., 2012, Kodama et al., 2017] and appropriate terminology could prevent children from expressing the models they have in their heads [Denham, 1993, Thatcher and Greyling, 1998]. As a consequence, elicited mental models may be incomplete or noncompliant with the actual model. Moreover, researchers criticise that interviewing might cause panic in children due to the test-like atmosphere it conveys [Denham, 1993, Thatcher and Greyling, 1998]. This may be ethically critical and it might threaten the validity of the results. Also, since with an expected number of participants of around 70, we must take the time factor into account. Techniques that can be applied to several participants in parallel are therefore more suitable for our study.

Using drawings is also a popular technique, but its effectiveness strongly depends on the level of abstraction of the matter

Drawing: Another widely used technique is drawing, which potentially offers a solution to some of these challenges [Doyle et al., 2022]. Drawing is especially popular in research with children. While it avoids stress for the children and saves time by allowing the elicitation from multiple participants at once, this technique also facilitates the actual creation of mental models through the process of drawing [Glynn, 1997, Kodama et al., 2017]. One of the first works to use this was presented by Denham [1993]. To elicit children's mental models of computers, they asked their young participants to draw what the inside of a computer looked like under the lid. Similarly, Pancratz and Diethelm [2020] investigated children's conceptions of computing systems architecture by asking them to draw their ideas of what smartphones, video gaming consoles, and robotic vacuum cleaners look like from the inside. What is noticeable is that both works examine constructs of concrete substantial matter that can only be either right or wrong and can therefore be directly elicited and evaluated through drawing. This, however, does not apply to dark patterns, which are abstract, imperceptible constructs that cannot be drawn [Brodsky et al., 2021] and must therefore be elicited indirectly [Doyle et al., 2022, Marhan et al., 2012, Jones et al., 2011]. The similarly abstract topic of privacy was investigated by Oates et al. [2018], who gave participants the prompt to draw what privacy meant to them. This instruction allowed symbolic and metaphorical, as well as analogical drawings, which require appropriate methods for a valid analysis.

Drawing can be either used as a direct or indirect elicitation technique

However, Kodama et al. [2017] argue that the precise formulation of the instruction in drawing tasks has a notable effect on the expression of the child's mental model. The instruction must therefore be carefully designed, depending on the nature of the concept under investigation and its level of abstraction. For example, various projects over the years have explored (children's) mental models of the internet from slightly different viewpoints. Depending on the researcher's point of view, prompts for the same research question can range from very general (e.g., 'Draw a picture to show me what the Internet looks like' [Brodsky et al., 2021] and 'Draw a picture of your perceptions about the Web' [Dinet and Kitajima, 2011]) to very specific (e.g., 'Explain how the internet works and draw a general diagram of it' [Kang et al., 2015] and 'Draw or sketch how you think the Internet is structured and how it works (in a way that a stranger would recognise it)' [Thatcher and Greyling, 1998]). Another challenge of the drawing technique lies in the scientific evaluation of the resulting drawings; some children lack the manual dexterity skills necessary to depict what they want to express [Marhan et al., 2012] and others express themselves only in symbolic and inscrutable ways [Panagiotaki et al., 2006]. As a result, drawings are sometimes up to the researcher's interpretation. Therefore,

Instructions need to be especially carefully formulated in drawing tasks

The interpretation of the drawings is sometimes up to the researcher

Pridmore and Lansdown [1997] suggest asking participants to include textual annotations explaining their drawings to avoid misinterpretations and false deductions.

Besides interviewing and drawing tasks, there are a number of other less popular elicitation techniques, some of which we will briefly cover in the following. However, most techniques are designed for more concrete questions or contexts and are therefore not always applicable.

Recognition tasks, arranging cards, and diagramming are also effective elicitation techniques

Recognition: Criticising that drawing and open questions in interviews would only reveal children's naïve mental models, which would be unsuitable for scientific constructs, Panagiotaki et al. [2006] proposed to use recognition tasks instead for more scientifically accurate responses. As an example, they explored children's representations of the earth by asking them to select a shape that most closely resembled the earth's shape from a given set of 3D models. They discovered that children knew more about the earth than prior work, using free-recall techniques, had suggested. This, again, underlines the importance of using appropriate methods and instructions to ensure validity.

Arrange Cards & Diagramming: Another technique that uses recognition is *Arrange Cards*, where participants are tasked to spatially arrange pieces of concepts written on a set of cards in a way that matches their representation of a construct. Similarly to the diagramming approach [Kang et al., 2015], this method was shown to be effective for eliciting children's internal structures of knowledge of a particular topic [Marhan et al., 2012, Jones et al., 2011].

There are even more elicitation techniques reported in the literature

Partcipatory design techniques & others: A relatively new approach to mental model elicitation of children is to apply participatory design techniques to assess complex and hard-to-express representations. For instance, Yip et al. [2019] used a participatory design technique called *Line Judging* to extract a conceptual model of children's perspectives of 'creepy' technologies. The method allows researchers to evaluate the feelings and attitudes of children in a comparable but more playful way to semantic differential scales. Further techniques reported in the literature include using metaphor techniques [Gentner and Gentner, 2014, Bostrom, 2008], laddering [Doyle et al., 2022], and Repertory Grids [Ben-Zvi Assaraf et al., 2012].

To conclude, as the literature suggests, none of the elicitation techniques comes without challenges and choosing the right method is often a trade-off, depending on the specific research question [Doyle and Ford, 1998]. However, one way to counteract the disadvantages of different techniques is the triangulation of multiple techniques [Grenier and Dudzinska-Przesmitzki, 2015]. *Teach-back* is a technique which implements this by letting participants explain their drawings, diagrams, textual responses etc. to each other or the researchers [Marhan et al., 2012]. For example, Clift et al. [2022] and Kodama et al. [2017] both combined drawing and explanation by first asking their participants to draw images according to a given prompt and then, in a discussion part, have them explain their drawings to the researchers to provide a broader perspective on the results. Nevertheless, researchers should always be cognisant of the inevitable trade-offs and that, as Norman [1983] put: 'we must discard our hopes of finding neat, elegant mental models, but instead learn to understand the messy, sloppy, incomplete, and indistinct structures that people actually have'.

Triangulation of different techniques is a way to deal with the trade-off of each individual technique

3.1.2 Questionnaire Creation

After getting an overview of existing methods and brainstorming with three fellow researchers, the next step was translating the new findings into concrete ideas for our study design. In the following sections, we provide a detailed description of the general considerations we made in advance and the creation of the questionnaire per task. The complete, original questionnaire in German and an English translation can be found in the Appendices B.1 "Mental Models Questionnaire (German)" and B.2 "Mental Models Questionnaire (English)", respectively.

General Considerations

We designed the questionnaire in close dialogue with pedagogic professionals from the school

The first decision we made was to conduct the study through questionnaires, which would allow several people to participate in parallel and thus increase the number of participants. During the creation of the questionnaire, we were in close dialogue with the school social worker and the class teachers of the participating school classes to incorporate expertise in working with children. This also gave us the opportunity to ask questions about our participants in advance — without compromising the anonymity of the children — which was necessary to make certain design decisions. For example, we were informed that all of the children had good German language skills and that no child was known to have any red-green colour blindness. Therefore, no special considerations had to be made in this regard.

Given the available evidence on children and their cognitive abilities, which suggests that children have comparatively lower attention spans, we decided against using well-established but complex scales such as Chaouachi and Rached [2012]'s two-dimensional scale to measure perceived deception in advertising. Instead, we decided to minimise cognitive effort by forming the questionnaire in a playful way, as proposed by Mertala [2021]. This has the advantage that children exhibit deeper involvement and superior performance under playful conditions than under formal conditions because they have more confidence and less pressure of never 'getting it wrong' [Mcinnes et al., 2009].

The questionnaire should be playful and contextualised in a guiding story

Furthermore, previous work suggests that including small, introductory stories with personal references in the questionnaire makes it easier for them to process information in the study tasks [Bell, 2007]. Therefore, we decided to contextualise the tasks within a personal guiding story that carries through the questionnaire. More concretely, the children were asked to imagine that they were players of the fictional card game '4 Jagd' (German, '4 Hunts'). In this game, players receive four playing cards with slightly different screenshots of an app, on each turn. Additionally, each player is given a secret goal, which they must try to achieve by selecting the one card that matches the goal best. If they succeed, they win the round. As the literature suggests that children tend to be particular about text and instructions [Bell, 2007], we intentionally did not specify the game in more detail to not confuse our participants with potentially inconsistent game rules. Also, we only revealed as much detail about the game as necessary for the task at hand. This was done to avoid overwhelming the children with long texts and initially unnecessary information. Thus, the rules of the game unfolded gradually over the course of the study.

Along with the guiding story, we tried to base the context of the tasks on what is familiar to children in order to create more realism and generalisability. Accordingly, we considered only mobile scenarios due to their prevalent role in children's social lives [Rohleder, 2022, Di Geronimo et al., 2020], and dark patterns that children might have encountered already, e.g., dark patterns that are common in mobile games, but no shopping-related patterns, such as *Sneak into Basket*[1]. However, when selecting the dark patterns that we wanted to include in the study, we had to take into account that not all common patterns were suitable: since in our questionnaire we could only provide static screenshots with no interaction possibilities, we excluded a number of prevalent but statically not representable patterns from our selection. For instance, *Nagging*, where the user is persistently interrupted by recurring requests [Brignull et al., 2010], was reported among the most prevalent dark patterns in gaming apps [Lupiáñez-Villanueva et al., 2022] but its essence cannot be illustrated by a screenshot and it is, therefore, unsuitable for our study.

We used only familiar scenarios, i.e. mobile context and dark patterns the children have likely seen before

For the design of the individual tasks, we decided to follow a similar process to Yao et al. [2017]'s by looking at our question from different perspectives and using different methodologies. Drawing from our review of elicitation techniques in Section 3.1.1 "Existing Mental Model Elicitation Techniques"), we used multi-method mental model elicitation [Grenier and Dudzinska-Przesmitzki,

We adopted a multi-method mental model elicitation approach

[1] `https://www.deceptive.design/types/sneaking` *Accessed: May 10, 2023*

2015]: overall, our methods combine recognition and free-recall tasks, as well as drawings and discussions. For most tasks, we also asked the participants to provide textual justifications for their responses. This could be an effective way to increase validity by preventing uncertainties and misunderstandings, as Pridmore and Lansdown [1997] suggest. In the exact formulation of the tasks, we had to be mindful of asking the 'right' questions and, above all, using simple, unambiguous sentences, as children are particularly susceptible to misunderstanding tasks, for example, because they take them too literally [Bell, 2007, Prokop et al., 2008]. We also avoided the term 'task' to avoid creating a test-like atmosphere and instead used the word 'part' to refer to different sections of the questionnaire. Eventually, after a first draft of the questionnaire was created, we tested the entire procedure in a pilot study with a small group of children of the target age group. This was shown to be particularly important when investigating populations that are as unexplored as children [Bell, 2007]. The identified weaknesses, such as ambiguities in the wording of the tasks, were adjusted before we conducted the study in the classrooms.

We conducted a pilot study to identify weaknesses of the questionnaire

In the following, we outline our specific considerations for the design of each individual part of the questionnaire.

Informed Consent & Demographics

The first page of the questionnaire includes a consent form for the participating children, as well as questions about participant demographics. Beforehand, the parents or legal guardians had already signed a comprehensive consent form with information on the procedure and purpose of the study, as well as the anonymous processing of their children's data. The consent form for the children themselves was given in simplified language. Therein, they were informed about the duration and procedure of the study and it was explained that they could learn more about the topic of 'media literacy' in the process. The exact topic of the study was not specified further, similar to the declaration of consent sent to the parents or legal guardians, in order

Both children and their parents/guardians received consent forms to confirm participation to avoid any bias. The children were also assured that all information would remain anonymous. Moreover, it was made clear that participation was voluntary and that they were free to stop or ask for a break at any time. Finally, they were invited to tick a check box confirming that they had read the information and wished to participate in the study if this was the case. Only those participants who had both ticked the check box and submitted a parent or legal guardian's signature were considered in the analysis of the data.

The second part included questions on personal information to gather demographic background details about our participants. It comprised questions about the age and gender of the child, each to be answered in a free text field. In addition, to understand their mobile device use, we asked them to indicate how often they approximately used a smartphone or tablet by ticking one out of five response options ranging from 'not at all' to 'at least once a day'. Finally, they were asked to specify which apps they thought they used most often, also in a free text field. It is notable that, when designing the demographics section, we were aware of the limited reliability of the self-reported data, especially for the last question. Nevertheless, we decided to keep this form, since approximate background data and assessments were sufficient for our purpose and we wanted to keep the questionnaire as simple as possible.

In the second part, we collected demographics data

Part 1 — Spontaneous Judgements of Websites Using Dark Patterns

Before we started the actual mental model elicitation, the aim of the first part of the questionnaire was to give the children the opportunity to form mental models of dark patterns by exposing them to different types and degrees [Grenier and Dudzinska-Przesmitzki, 2015] of dark patterns. More than that, we wanted to get an impression of how children spontaneously judged websites using varying degrees of such malicious designs. We decided to examine this under the three dimensions *visual appeal*, *perceived usability*, and *trustworthiness*, since Lindgaard et al. [2011] dis-

We measured children's spontaneous judgements of websites using varying degrees of dark patterns

Figure 3.1: The four designs children were asked to judge in Part 1. They comprise different quantities and intensities of dark patterns: *U5* is dark pattern-free. *N2* contains one mild dark pattern called Confirmshaming. *G6* used one aggressive Visual Interference pattern. *J9* combines the two previous dark patterns Confirmshaming and Visual Interference.

covered that those were the three essential factors of different cognitive demands to judge websites.

In line with the guiding story, Part 1 introduced the first game round and the children received their first four playing cards (see Fig. 3.1). All cards show screenshots of the same fictional mobile game where all three lives have been used up. To get new lives and continue playing, the user must either wait a specified time or, as advised by a prompt, watch an ad. The screenshots differ slightly in the combinations of quantity and intensity of the dark patterns they use. As previously mentioned, we only included those pattern types that our participants could potentially have seen before. We based our designs on Luguri and Strahilevitz

[2021]'s scheme of differentiating dark patterns by intensity into mild and aggressive patterns and combined them with the number of occurrences from zero to two to create the following designs:

1. The screenshot labelled 'U5' is dark pattern-free. It shows two visually equal buttons, with the neutrally labelled options 'Ok, watch ad' or 'Don't watch ad', respectively. Both options seem equally tempting.

The designs comprise one dark pattern-free design...

2. Screenshot 'N2' contains one mild dark pattern, being *Confirmshaming*: while the two buttons are, again, visually equally designed, the label on the left responds to the prompt to watch an ad with 'Ok, I want to keep playing!' and the right one is labelled 'I want to wait for new lives and not continue playing'. In this scenario, users might feel pushed to click on the left button as they want to keep playing.

...one using a mild Confirmshaming dark pattern...

3. Screenshot 'G6' shows one aggressive dark pattern named *Visual Interference*: unlike the two previous designs, this one has only one button, which is labelled 'Watch ad', with no obvious option to reject. The close button, which is the only alternative to click on, is a fine cross icon in the upper left corner of the window. In contrast to the previous screenshots, where it was designed as a clearly visible red close button in the upper right corner of the window, the button on this image is barely visible at first glance. Users might miss this option and feel forced to click on the supposedly only option which makes them watch an ad.

... one with an aggressive Visual Interference ...

4. Screenshot 'J9' combines the mild *Confirmshaming* dark pattern from N2 and the aggressive *Visual Interference* from G6: the close button is still imperceptible and there is only one clearly visible button, which is labelled 'Ok' and accepts to watch an ad. The reject option is designed borderless and with a transparent background. Hence, it is hardly recognisable as a clickable button. Its label reads the same Confirmshaming-formulation as the right button from screenshot N2, urging users to not select this

... and one combing both patterns

option, provided that they have noticed it in the first place.

Children judged the designs according to how beautiful, complex, and trustworthy they perceived them

After receiving the four screenshots, the children were asked to indicate how they perceived each, regarding Lindgaard et al. [2011]'s three dimensions *visual appeal*, *perceived usability*, and *trustworthiness*. To measure this, we needed a child-friendly but expressive method. Inspired by the previously mentioned methods *Line Judging* [Yip et al., 2019] and *Arranging Cards* [Marhan et al., 2012, Jones et al., 2011] (see Section 3.1.1 "Existing Mental Model Elicitation Techniques"), we developed a hybrid approach that captures both the participants' impressions on a semantic differentials scale and their relative rankings from spatial rearrangements of the screenshots: for each of the three attributes, the questionnaire included a ruler-like horizontal line with 21 equally distributed tick marks. The leftmost, centre and rightmost tick marks were highlighted and labelled with extreme negative, neutral, and extreme positive verbal labels, depending on the corresponding dimension they referred to. We used verbal labels, as children were shown to understand them more easily than numeric ones [Bell, 2007]. However, we simplified the previously specified terms to more familiar and comprehensible ones for children, namely *beauty* (i.e., not beautiful/neutral/very beautiful), *complexity* (i.e., not complex/neutral/very complex), and *trustworthiness* (i.e., not trustworthy/neutral/very trustworthy).

We developed a hybrid scale, combining semantic differentials and ranking

The task was to position each of the four images on each of the three lines where it most closely corresponded to the child's impression by marking the respective tick mark with a cross and the label of the screenshot. To demonstrate how to proceed with the task and special edge cases, we used an unrelated example to show the class how we would arrange various pictures of animals according to their cuteness on a cuteness scale. Untypical for a conventional ranking, they were explicitly allowed to put several images in the same position, since we did not want to force them to make a decision which could both create a stressful situation and potentially suggest preferences that are not accurate. The decision to use 21 tick marks was the result of a compromise between the sensitivity of the scale and its analytical opportunities. Using continuous scales or those that have a high number of tick marks considerably increases the granularity and discrimination of participants' responses, which helps to reduce ceiling and floor effects [Chyung et al., 2018]. However, these scales pose a challenge in the evaluation and analysis of the responses: it is not only more difficult to interpret the position of a handwritten marking on a (nearly) continuous scale, especially with child participants, but the number of opportunities for statistical testing is also reduced [Chyung et al., 2018, Lazar et al., 2017]. Additionally, providing too many tick marks might overwhelm and stress out participants. On the other hand, discrete scales with a small number of tick marks are easy to interpret and analyse, and provide more analytical opportunities, while being less expressive and granular than continuous scales. Also, providing just a small set of predefined response options might discourage elaboration [Lazar et al., 2017, Chyung et al., 2018]. Thus, as a compromise, we used a discrete scale with a relatively high number of tick marks to approximate a continuous scale, which combines the advantages of both scale types.

We suspected design labels and orderings to be potentially confounding, so we attempted to counteract them

Finally, we had to counteract potential confounding factors that could influence the positioning of the screenshots on the scales. On the one hand, we developed a labelling system to distinctly identify the four screenshots without introducing bias due to the nature of the label. For example, while we wanted to keep the cognitive effort low by using short labels, using individual letters or numbers may inherently convey an ordinal meaning which might influence the participants. We, therefore, used different two-character labels, composed of one letter and one number, each. We expect that this helps keep participants focused on the features of the designs themselves, rather than being influenced by preconceived conventions regarding specific letters or numbers. Second, we considered potential order effects through the order in which we presented the four screenshots in the questionnaire. To counteract these, we applied a balanced 4x4 Latin square, which, ultimately, yielded four different versions of the first part of the questionnaire.

Part 2.1 & 2.2 — Linking of Specific Design Elements With Manipulations

The research on children's mental models of dark patterns revolves around the question of whether they understand the influential nature of certain design elements, such as colours, shapes, wording, etc. More specifically, we wanted to investigate whether children are aware that such designs are used strategically to pursue certain (malicious) intentions, and if so, what those intentions are. These questions can be looked at from different angles.

Children were asked to select a design that would achieve a given manipulative goal

On page 3 of the questionnaire (see Section B "Mental Models Questionnaire"), we investigated whether children can select the one design from a choice that is most likely to achieve a given manipulation goal. For this, we adopted Panagiotaki et al. [2006]'s model selection approach for mental model elicitation. This technique is less cognitively demanding than free-recall questions and is, thus, well suited to provide a less overwhelming starting question before transitioning to more complex open tasks (see Section 3.1.1 "Existing Mental Model Elicitation Techniques"). In this part, the children were again shown two of the screenshots from Part 1 (see Section 3.1.2 "Part 1 — Spontaneous Judgements of Websites Using Dark Patterns") and, in line with the game '4 Jagd' from the guiding story, a goal they had to achieve to win the round: they were asked to select the screenshot that would get the most people to watch an ad. The two screenshots to choose from were image U5, which is dark pattern-free, and image G6, which uses the aggressive dark pattern Visual Interference to urge users to watch the ad (see Fig. 3.1). We chose these two screenshots as we wanted to include the dark pattern-free baseline (U5) and an option that contained exactly one visual dark pattern (G6) so that we could attribute decisions to the mere existence of dark patterns without mixing too many factors, e.g. by including designs that differ by more than one dark pattern, like J9. The children were asked to tick the corresponding check box for their selection and to give a short textual justification afterwards. This would help us to understand whether their decision was based on the Visual Interference pattern, whether there was another motive, or

They received one dark pattern-free design and one containing one Visual Interference whether the selection was probably arbitrary.

To enable us to draw further conclusions about their motives, we preceded this question with a smaller task: before the children were to make their selection in Part 2.2, they were first asked to mark and textually describe any differences they could find between the two screenshots, similar to a spot-the-difference puzzle. For example, in the case of elusive justifications for the selection, this would allow us to check whether it was potentially due to the child not recognising the differences between the two designs in the first place, which could indicate that the choice was made arbitrarily. Another reason why we included this task was to provide a starting point for thinking about the selection for the next task, so they would not be overwhelmed by it. While we are aware that this task can strongly influence the choice in Part 2.2, the pilot study showed that it was necessary in order to make the task clearer for the children to understand. Still, we will take this potential bias into account in the evaluation of the results.

Before choosing, they marked and described differences between the designs

Part 2.3 — Identifying Intents Behind Design Elements

While we had previously approached the question through recognition methods, we wanted to triangulate by adding an indirect open-question technique that could open up new perspectives [Panagiotaki et al., 2006, Marhan et al., 2012]. Part 2.3 of the questionnaire aimed at understanding whether children would realise the underlying (malicious) intention behind including a dark pattern in a design. To explore this question, we reversed the task from Part 2.2: for a given design containing dark patterns, they were asked to identify the underlying manipulative goal. More precisely, the children again received two screenshots (see Fig. 3.2), which only differed in one using a dark pattern and the other being fair. Following our guiding story, we explained that it was now another co-player's turn in the game '4 Jagd' and that they chose image T3 from the selection between the two screenshots to try to achieve their secret goal. In a text box, the children were asked to describe their presumption of what secret goal this could have

Children were asked to identify the intent behind choosing a design containing Visual Interference

Figure 3.2: The two designs of a cookie consent the children were given in Part 2.3. It contains one accept (left) and one reject button (right). *F7* is dark pattern-free, while *T3* entails a Visual Interference as the reject button is greyed out.

been and to justify their answer.

Two new screenshots were used for this task, showing a simple cookie consent. While both screenshots inform users that the site uses cookies 'for the best experience', which can be either accepted or rejected via a button, they differ in the visual appearance of the two buttons:

The task contained one dark pattern-free design and one with Visual Interference

1. Screenshot 'F7' was designed to be dark pattern-free and fair by showing two buttons of the same colour, size, and font, labelled 'Accept all' and 'Reject all', respectively. Both buttons are equally tempting.
2. Screenshot 'T3' uses, again, a *Visual Interference* dark pattern. While the 'Accept all' button is clearly visible because of its dark colour, which contrasts strongly with the light background and its font colour, the reject button is comparatively easy to miss; the button itself and its font are similarly light-grey coloured as the background and the button and font size are considerably smaller as compared to the accept option. The option to reject all cookies is, thus, not only easy to miss but it could also be misinterpreted as being inactive. Hence, it is more likely and intended that users would accept all cookies.

While the first screenshot F7 was, again, designed as a baseline, we chose this specific manipulative design for the second screenshot T3, as it is highly prevalent on the internet and children will likely have encountered it already. This might more accurately reflect what is already in children's minds instead of having them create completely new mental models of a construct they have never seen before.

To maintain consistency, we used the same labelling system with the two-digit letter-number combinations as in previous tasks to address different screenshots. However, we did not use a Latin square in this task because there was no risk of bias arising from order effects. We are, again, aware that previous tasks might influence the results of this task but, as mentioned before, this task order was shown to be necessary in a pilot study in order to clarify the instruction of Part 2.3. We will, however, consider this limitation in our evaluation.

Previous tasks might influence children's responses for this part

Bonus Part — Putting Known Manipulative Tricks Into Practice

In the next task, on page 5 of the questionnaire, we approached the research question again from a new perspective, using a drawing technique to make the children draw dark patterns themselves [Morrison et al., 2021, Jones et al., 2011]. Building on Part 2.3, we asked the children how the goal they had identified in the previous task could still be achieved by changing individual design elements of the cookie consent from design F7, other than the *Visual Interference* they were given in image T3. They could draw their ideas for an alternative design in a box and were asked to additionally describe their design textually. On the one hand, this task was intended as a transfer task from what children already knew. On the other hand, it was aimed at recognising whether children have an understanding of the various possibilities that exist for online manipulation. What is particularly interesting about the task is the chance to investigate whether children themselves would use common dark patterns that they have encountered in everyday life, which would allow deductions to be made about whether manipulative patterns have probably been

In a bonus task, children redesigned the fair design in a way to manipulate users to accept cookies *implanted* in children's minds through their prevalence.

The task was voluntary to not stress children with its potentially high complexity

For this particular task, we decided to use the drawing technique because we expected that its free-recall nature might make it difficult for children to articulate their ideas textually. Nevertheless, we decided to ask for an additional textual explanation, both to avoid misinterpretation of the drawings (see Section 3.1.1 "Existing Mental Model Elicitation Techniques"), but also to give the children the complementary opportunity to express those ideas that are easier to describe textually. We regarded this task to be comparatively difficult, which was bolstered in the pilot study. There was, thus, a risk of children feeling like they had failed if they could not answer it, for example, because they had not been able to identify a goal in the previous task or because they just could not come up with an idea. In Section 3.1.1 "Studies with Children", we have already discussed that this threat is particularly present in studies with children and that the study should be designed accordingly to mitigate this. Therefore, we declared this entire task as a voluntary bonus part, to not create any pressure.

Part 3 — Recognition of Dark Patterns

In Part 3, children marked all manipulations they could find on a screenshot

After a short verbal briefing about online manipulation based on the previous tasks, we finally explored on the last page the question of what children understood by the term 'online manipulation'. More specifically, they were given a screenshot in which we incorporated four dark patterns and asked our participants to circle all the manipulations they could recognise in the image and, as before, to give textual reasons for their selection.

The screenshot for Part 3 contains four dark patterns:
1. Toying with Emotions...

The screenshot we designed for this task (see Fig. 3.3) represents the same app as in Part 1 and shows a state in which the user wants to close the app and is asked if they are sure they want to quit the game. There are two buttons available: a red button labelled 'Ok' that closes the app, and a wider green button labelled 'Cancel' that takes the user back to the game. This use of the red and green colour conventions for the two buttons is a Toying with Emotions

Figure 3.3: The design for the recognition task in Part 3 contains four dark patterns: 1) Toying with Emotions: the user is asked if they are sure they want to leave the app, with one 'Ok' button in red (left) and one 'Cancel' button in green (right). 2) Preselection: the checkbox for sharing the score with friends is preselected. 3) Confirmshaming: the formulation for the option to reject sharing the score is uninviting to select. 4) Trick Question: the cookie banner at the bottom of the screen uses double negations.

dark pattern (although with the differing sizes, it also contains Visual Interference elements), where users are supposed to feel drawn to the large green button, causing them not to close the app after all.

Moreover, the user can tick whether they want to share their score with their friends or not. At this position, two dark patterns are used: first, the option 'share score with friends' is preselected, which makes sharing data the default option if the user misses the question or ignores it due to time reasons. This corresponds to Gray et al. [2018]'s Preselection pattern. On the other hand, the alternative option

2. Preselection

is worded in a way that users may be reluctant to select it ('I play too badly to share my score with friends'). This corresponds to the *Confirmshaming* pattern and is, again, intended to tempt the user to share their score with friends.

3. Confirmshaming

Lastly, at the bottom of the page, there is a cookie consent, with the two options 'No' and 'Ok'. The *Trick Question* pattern is used here, as the wording is intentionally misleading and uses double negations ('If you don't want us to not save cookies, press "No"'), which results in the user having to select 'Ok' to reject the use of cookies and 'No' to refuse, contrary to what is intuitively and conventionally assumed.

4. Trick Question

We chose to conclude the questionnaire with another recognition task as this served well for this perspective of the question and also provided a cognitively low-stress closure after a long and probably exhaustive study. In order to obtain carefully weighed answers, we decided to include the number of hidden manipulations in the task description.

3.1.3 Recruitment of Participants

Parallel to the creation of the questionnaire, we had to start recruiting participants well in advance. In the following, we will provide a brief overview of our target group and key considerations in the recruitment of participants.

We specifically targeted a sample of 10–11-year-old children: according to statistics about children in Germany by Rohleder [2022], this is the age when children take a technological leap and begin to use the internet more often. At the same time, parental control decreases. More specifically, the researchers estimated that in 2022, more than 90% of German children aged 10–12 used smartphones or tablets 'at least occasionally', with an average usage time of more than one hour per day. 86% even owned their own smartphone (see Fig. 3.4).

We targeted 10–11-year-olds because, at age 10, children take a technological leap

In order to reach larger sample sizes more easily and efficiently, we decided to conduct the study with entire school classes in the classroom. A school social worker acted as

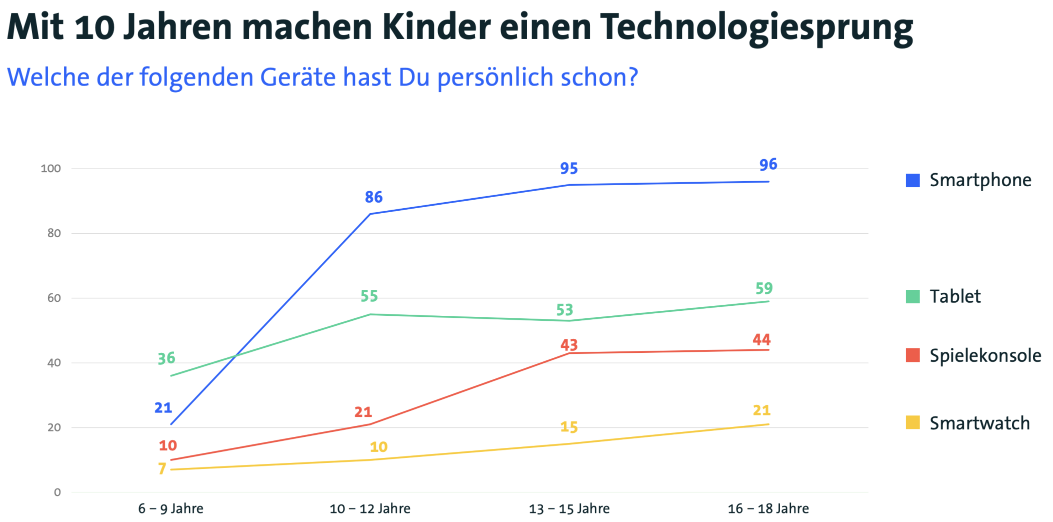

Figure 3.4: The diagram visualises the percentage of children (y-axis) per age group (x-axis) owing a smartphone (blue), tablet (green), game console (red), or smartwatch (yellow). It shows that at 10, children take a technological leap. Graphic taken from Rohleder [2022].

a contact person for a local grammar school, allowing us to carry out the study in three classes, each with about 25 5th-graders 10–11 years old. In return, we offered the students a short learning unit on dark patterns and the associated dangers on the internet and in apps, as well as tips on how to best protect themselves against them. The concept of the unit was developed in consultation with the school social worker and the corresponding class teachers. Its content complemented the existing media literacy curriculum, which is specified by the *Medienkompetenzrahmen NRW*[2] but does not include education regarding dark patterns or other malicious designs on the internet.

We conducted the study with three school classes of 5th graders of a local grammar school and offered a learning unit on dark patterns

Prior to the study, consent forms were distributed to all parents or legal guardians of children in the respective classes. They were informed about the content and procedure of the study and that the children's data would be collected and analysed anonymously, so that no conclusions may be drawn to the child's identity. They were also assured that the data would be used for scientific purposes only. When

[2] `https://medienkompetenzrahmen.nrw/` *Accessed: May 10, 2023*

describing the content of the study, we kept the wording vague and only provided the necessary information to avoid bias in the data. For this work, only the data of those children who submitted a consent form signed by a parent or legal guardian was processed. Due to the time-limited scope of this work, we made the conscious decision to use three classes from the same school to avoid long delays in coordinating multiple schools. This also facilitated the control of extraneous variables. However, we are aware of potential biases stemming from this and consider them in our interpretation of the results.

Only data from children who submitted a signed consent form from parents or guardians was analysed

3.1.4 Study Setup & Procedure

After we had prepared the questionnaire and recruited participants, we conducted the study. In this section, we will briefly outline our study setup at the school and the procedure.

The 6-page questionnaire described in the previous sections was administered within two weeks to three 5th-grade classes of about 25 students each. The entire study took around 90 minutes. This time included a short break after the bonus drawing task, as well as two briefings and question & discussion rounds, comparable to Yao et al. [2017]'s study. The first round of discussion took place after the break, before the last task. We discussed the children's answers and thoughts on the previous tasks and, together, concluded from our discourse that one could trick or manipulate users into doing things they did not want to do, e.g., through the look of the website. This was necessary to introduce the last task in a meaningful way, where the children themselves had to find manipulations on the screenshot of an app. However, up to this point, we had actively avoided using the word 'manipulation' or equivalent, so as not to bias the children. Hence, with the briefing, we wanted to create a less confusing transition to Part 3, while educating and warning the children about the phenomenon of online manipulation. We held the second round of discussion after the last task to smoothly wrap up the study. Also, we used the dialogue to explicitly educate the chil-

We included question & discussion sessions to educate children about dark patterns dren about dark patterns and their threats, and how one could protect oneself from them in the future. In the end, we used the opportunity to answer questions and exchange experiences.

In order to set the pace and maintain control, we handed out the individual sheets of the questionnaire one after the other, after we had verbally explained the respective task and answered questions. This was also important to avoid children feeling that they had failed if they worked more slowly and could not finish the questionnaire in the time available, since, in this case, they would not be able to see how many tasks they had missed. Furthermore, we kept the verbal instructions short and clear to not influence the results [Kodama et al., 2017]. We also made sure to emphasise throughout the study that there would be no right or wrong answers, as pointed out in Section 3.1.1 "Studies with Children".

We explained and handed out the tasks one after the other to set the pace

We conducted the study together with one or two fellow researchers and the respective class or substitute teachers. This allowed us to conduct the study with the entire class at once, thereby saving time: whenever a question arose, we were able to take time for each child individually so that no one would feel overwhelmed or left alone. At the same time, we could make sure that someone was always keeping an eye on the children, which was necessary to control the risk of falsified results. For example, we had to take care that no loose questionnaire sheets were exchanged and that solutions were not shared or copied. Besides that, the role of the teachers during the study was important because they knew the children personally. As such, they served as a contact person the children could trust, thereby reducing potential anxiety and distress. They also had the pedagogical expertise to intervene in case of difficult situations, e.g. children feeling too much under pressure.

Several researchers and one teacher were always present at the same time to answer questions and for supervision

Also, the teachers helped us check the status of the parents' or guardians' signatures on the consent forms. To not exclude the children who had not submitted a signature, we allowed them to participate in the study as well but carefully marked the respective questionnaires with the help of the teachers. That way, we could put them aside afterwards

Children with missing signatures were not excluded from the study itself, but from the analysis and not consider them in the evaluation. If signatures were submitted later, we removed the mark with no residue and included the results in our evaluation.

Ultimately, we stapled all the pages per participant, collected the completed questionnaires and scanned them before we started to analyse the data.

3.1.5 Ethical Considerations

We followed ACM guidelines and child-specific principles for ethical conduct

When designing and conducting the study, there were a number of ethical considerations that we took into account. We primarily followed the *ACM Code of Ethics and Professional Conduct*[3], which provides guidelines for the ethical conduct of computing professionals. This includes but is not limited to the principles of always being honest, trustworthy, fair and never discriminatory in the conduct. In addition, one must adhere to the principle of justice and to national and international regulations.

For studies with minors, however, there are a number of ethical guidelines that must be respected eminently. For our work, we aimed to comply with the ethical standards established by the *'Ethical Research Involving Children'* project[4], which is an international, collaborative initiative founded by researchers, child advocacy organisations, and ethics committees, and focuses on upholding ethical standards in research involving children. It encompasses broadly the following principles:

1. principle: avoid harm

Avoid harm throughout or as a consequence of participation. For our study, we implemented strategies to reduce distress for our underage participants. In Section 3.1.1 "Studies with Children", we highlight possible sources of overload, stress, and discomfort for children as study participants. Accordingly, in Section 3.1.2 "General Considerations", we describe the appropriate measures we took to

[3] `https://www.acm.org/code-of-ethics` *Accessed: May 12, 2023*

[4] `https://childethics.com/ethical-guidance/` *Accessed: May 12, 2023*

prevent these. For example, we designed the questionnaire tasks to be playful, chose age-appropriate contexts, and ensured that a trusted teacher was always present during the study.

Respect privacy and honour confidentiality. We anonymised all responses from the questionnaire with no possibility of drawing conclusions on the child's identity. Also, we avoided collecting sensitive data from the children.

2. principle: privacy and confidentiality

Obtain informed consent from both the participants and their parents or legal guardians. The informing part contains all relevant information about the study, e.g. the purpose, risks, and confidentiality measures. In the consenting part, our participants and their parents could confirm their voluntary participation or withdraw. The informed consent is outlined in more detail in Section 3.1.2 "Informed Consent & Demographics".

3. principle: informed consent

Compensate the children for participating in the study. Our aim was to ensure a lasting impact on the children. Thus, we incorporated an educational component on dark patterns, allowing them to deepen their knowledge, foster their ability to safeguard themselves, and engage in discussions to share their experiences and ask questions.

4. principle: compensate participants

3.1.6 Participants

After excluding nine participants from the evaluation due to missing parents' or guardians' signatures on the consent form, we eventually reached a total of 66 participants from three different 5th classes. Their ages ranged from 10 to 12 years old ($M = 10.5$, $SD = 0.5$), with 51.5% 10-year-olds, 47% 11-year-olds. Only one child indicated to be 12 years old. 56.1% identified as female, 43.9% as male. Regarding their frequency of smartphone and tablet use, 68.2% reported using their devices at least once a day. 22.8% indicated using them several times a week, 4.6% at most once a week, and 3% did not use them at all.

Our 66 participants were aged 10–11 years old

A majority uses their mobile devices every day

When asked which apps they used most often, WhatsApp

(62.1% of participants) and YouTube (56.1%) were by far the most frequently mentioned. The most popular app categories were communication apps (86.4% of participants), e.g. WhatsApp and Microsoft Teams, and entertainment apps (69.7%), e.g. YouTube and TikTok. A more detailed overview of all apps mentioned, together with their frequency of mentions, a categorisation into app categories, and a list of dark patterns contained in each app, can be taken from Appendix C "Most Frequently Used Apps and Dark Patterns They Contain". More details about our procedure for categorising the apps and collecting the accompanying dark patterns are provided in Section 3.1.7 "Data Analysis". According to the school social worker and the class teachers, the children had already acquired basic knowledge in the field of media competence and internet dangers, e.g. cyber-grooming and cyberbullying, from the curriculum. However, dark patterns or other manipulative designs were not taught. We, therefore, expected little to no prior knowledge of dark patterns.

Communication and entertainment apps were most popular

3.1.7 Data Analysis

Because of the variety of task types we used in our questionnaire, we applied quantitative and qualitative methods in the data analysis. We first analysed each part of the questionnaire individually and drew cross-task conclusions at the end. For descriptive and inferential statistics of the quantitative data – mainly for demographics and Part 1 – we used tools such as *Microsoft Excel*[5] and *R*[6], which is a programming language for statistical computing. However, we analysed the majority of the data with qualitative methods, using a combination of thematic analysis [Braun and Clarke, 2006] and content analysis [Mayring, 2014]. While in content analysis, the data is quantified, for example by counting the occurrences of codes, and further analysed using quantitative-appropriate methods, thematic analysis follows the approach of identifying patterns or themes in the data to develop an in-depth understanding

We used quantitative and qualitative methods for the data analysis

[5] `https://www.microsoft.com/en-us/microsoft-365/excel` *Accessed: May 10, 2023*

[6] `https://www.r-project.org` *Accessed: May 10, 2023*

of the construct. We decided to use the inductive thematic analysis approach most of the time, as it is flexible, explorative, and thus appropriate for our broad research question [Braun and Clarke, 2006]. Furthermore, as no coding frameworks exist for our unexplored field, using deductive methods was not an option for most of the tasks.

During the coding and evaluation process, we roughly followed the steps proposed by Braun et al. [2019] for thematic analyses: (1) *Familiarisation* by reading and rereading the data, (2) *Generating codes* that reflect the meanings of each response, (3) *Constructing themes* by grouping codes into broader categories, (4) *Revising themes* by reviewing and adjusting themes, (5) *Defining themes* by interpreting the findings arising from themes and relationships between them, and (6) *Producing the report* by summarising findings, e.g. themes, patterns, and implications. In practice, most of these steps consisted of several iterations to check, adjust and interrelate the naming of codes and grouping into themes. We also discussed our codings with other researchers who were uninvolved in the topic to ensure higher reliability. For the entire qualitative analysis, we used the coding software *MAXQDA20*[7], because it offers various formats for organising the responses (e.g. per participant or per task), as well as possibilities for analysing and visualising the codes and themes. It was therefore a convenient and efficient tool to facilitate the entire process. In the end, we used the programming language *Python*[8] to generate further visualisations and to investigate cross-task relationships. In an attempt to reduce bias in the distribution of codes during our content analyses, we adjusted the number of participants for each task to include only those who had completed it. In the following paragraphs, we further outline our approach to analysing the individual parts of the questionnaire.

We applied thematic and content analysis methods to our qualitative data

Demographics: To understand our demographics, we first applied common descriptive statistics methods to the age, gender, and smartphone/tablet usage frequency of our participants. We coded the answers from the free text field for frequently used apps and classified the mentioned apps

[7] `https://www.maxqda.com` *Accessed: May 10, 2023*
[8] `https://www.python.org` *Accessed: May 10, 2023*

into self-created app categories, which we derived according to the basic qualitative analysis approach [Saldaña, 2013]. Other than Di Geronimo et al. [2020], who referred to existing app categories from the Google Play Store, we decided to take an inductive approach and created categories that were less fine-granular and broad, and, thus, better represented the range of apps our participants used. We then assessed the frequency of occurrences of the apps and the categories and assigned a list of dark patterns used in each mentioned app for later cross-task analyses. For this, we relied exclusively on the data sets provided by Lupiáñez-Villanueva et al. [2022] and Di Geronimo et al. [2020][9], who addressed this question for a large number of apps. We did not revalidate the information by checking the apps ourselves, since this is not the focus of this work and would exceed the scope.

We categorised the apps and identified a list of dark patterns each app contains according to Di Geronimo et al. [2020] and Lupiáñez-Villanueva et al. [2022]

Part 1 – Semantic Differentials & Rankings: For the analysis of Part 1 of the questionnaire, we looked at the participants' responses from two different directions and applied descriptive and inferential statistics methods to each: first, we treated the three scales for beauty, complexity and trustworthiness as semantic differential scales with absolute values. We numbered each of the 21 tick marks per scale from left to right with a number between -10 and 10 and then interpreted the responses numerically within this range. Second, we treated them as rankings of the four screenshots, yielding relative positions. We mapped the marking positions of the screenshots per scale to a rank between 1 and 4, i.e. the right-most marking was assigned the value 1, while the left-most marking was assigned the value 4. In case several screenshots were placed in the same position, they also received the same ranking value and the next number was skipped. For example, if a participant placed three screenshots in the same position on the far right and one screenshot on the far left, this would translate into three times the rank 1 and once the rank 4. To investigate the differences between the four screenshot variations and the effect of the intensity and occurrences of dark patterns, we also applied inferential statistics to the absolute values from the semantic differentials scales, using *R*. Since the normal-

We quantitatively analysed the judgements for Part 1, treating them as both semantic differential and ranking scores

[9] Direct link to the data set: `https://zenodo.org/record/3601501` *Accessed: May 10, 2023*

ity assumption did not hold, we used the non-parametric Friedman rank sum test, along with Dunn's test with Bonferroni correction for pairwise comparison. It is important to note that the Friedman test only tests whether there are differences in the distribution of ranks and would therefore give the same results for semantic differentials and ranking scores.

We used Friedman rank sum test with Dunn's test with Bonferroni correction

Part 2.1 & Part 2.2 – Screenshot Selection: To assess whether the children were able to detect the differences between the fair screenshot U5 and the Visual Interference screenshot G6, we first quantified the participants' responses. We defined a boolean participant variable which we set to True if the participant had correctly marked and described the difference between the two images. Otherwise, we set it to False. Notably, we attempted to reduce threats to validity by also setting the variable to False when differences were correctly marked but explanations did not match or were missing. Moreover, to identify explanations for potentially wrong answers, we inductively coded the textual descriptions. In Part 2.2, children were asked to select which of these designs would deceive more users to watch an ad. Again, we created a boolean variable which evaluated True if children chose option G6 with the Visual Interference and False for the fair option U5. Next, we coded and analysed the textual explanations in accordance with the thematic analysis steps described earlier. In our analysis, we also looked at connections to the previous control task, Part 2.1.

We introduced participant variables to quantify the data and coded textual justifications

Part 2.3 – Goal Identification: As we wanted to test whether children had an intuition for what influence certain design elements might have on user behaviour, we, again, used thematic analysis for the responses to this part. In several iterations, we coded participants' speculations about the goal that was to be achieved by choosing screenshot T3 with the Visual Interference over the fair version F7. While coding, we divided the responses into two parts: 1) the goal they had identified, and 2) the justification for the goal. This resulted in two separate sets of codes. As before, we also used simple visualisation and computational tools to identify relationships between codes within the task and with findings from previous tasks.

We coded goals and justifications separately

Bonus Part – Drawing of Alternative Designs: In this transfer task, we wanted to investigate a higher level of children's understanding of dark patterns and online manipulation by tasking them to apply such techniques, themselves. We coded the resulting drawings with a focus on the manipulative elements applied, this time primarily using a deductive coding approach: wherever applicable, we used codes from an a priori set of codes we derived from existing dark pattern names, e.g. from Brignull et al. [2010], Gray et al. [2018], and Mathur et al. [2019]. For all other applied manipulative methods which we could not match with a known dark pattern, we inductively generated new codes, following the thematic analysis approach. Since we were also interested in whether the prevalence of dark patterns had already become so normal for children that they could instinctively use them on their own, we finally examined the resulting code system for overlaps with the list of dark patterns that we had previously identified from the frequently used apps in the demographics part.

We mixed inductive and deductive coding techniques to obtain a set of manipulative techniques from the drawings

We searched for overlaps of dark patterns used in the drawings and the list of patterns they had seen before in apps

Part 3 – Finding the Manipulations: Similar to the deductive coding in the analysis of the preceding drawing task, we mainly relied on a coding scheme consisting of names of dark patterns contained in the screenshot. While we defined this coding scheme beforehand during designing the screenshot, we inductively extended the codes over time when encountering alternative responses during the analysis. Then, we inductively coded the textual justifications and clustered the participant responses to identify frequent combinations of dark patterns found. For the sake of validity, we again ignored all responses in which the justifications for the markings were missing. As usual, we adjusted the superset of participants accordingly for numerical analyses.

We mainly deductively coded the identified manipulations for Part 3

3.2 Results

In the following sections, we report the results of the qualitative and quantitative analyses per task, as described before. When referencing individual participants, we refer to them anonymously according to the convention *Pxx*,

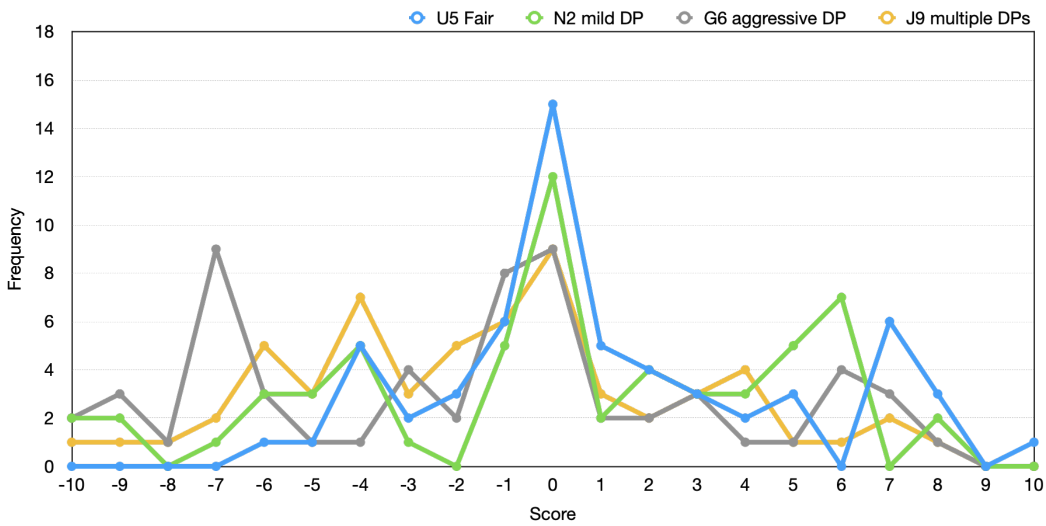

Figure 3.5: The line chart shows the distribution of the beauty scores per design for the semantic differential scales in Part 1 (N = 66). The y-axis depicts the frequency of each score which ranges from -10 (not beautiful) to 10 (very beautiful). Each line indicates one of the four designs U5, N2, G6, and J9.

where *xx* denotes the participant ID. A codebook giving an overview of all codes and categories used during the qualitative analyses is provided in Appendix D "Codebook".

3.2.1 Part 1 — Ranking & Semantic Differentials

Analysing the participants' responses to the semantic differentials scales with scores between -10 and 10 and the rankings, we obtained clear differences between the ratings of the four screenshots under the three aspects of beauty, complexity, and trustworthiness. Fig. 3.5 shows the distribution of the beauty scores for the four screenshots in a line chart, while Fig. 3.6 depicts in a stacked bar chart how often each beauty rank was assigned to the four images. All ranks 1 to 4 were covered for all screenshots on the beauty scale. The dark pattern-free, fair screenshot U5 achieved the highest beauty scores (M = 1.2, SD = 3.8, Mdn = 0.0, $Mode$ = 0), ranging from -6 to 10. Hence, it achieved the highest average ranking (M = 1.8, SD = 0.9, Mdn = 2.0, $Mode$ = 0). The mean of screenshot N2, using the mild Confirmshaming dark pattern, spread around the neutral cen-

The dark pattern-free design U5 was rated as the most beautiful, the Visual Interference design G6 was rated the least

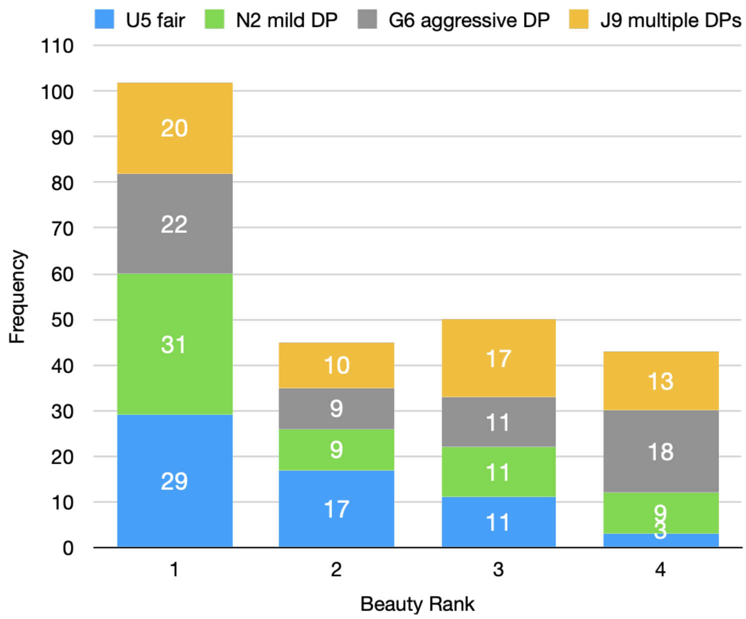

Figure 3.6: The Frequencies of beauty rankings for designs U5, N2, G6, and J9 (N = 66) in a stacked bar chart. It shows the distribution of rankings from 1 (most beautiful) to 4 (least beautiful) for each design, represented by different colours.

tre (M = 0.1, SD = 4.7, Mdn = 0.0, $Mode$ = 0), within the range of -10 to 8. It was on average ranked the second-highest (M = 2.0, SD = 1.2, Mdn = 1.0, $Mode$ = 0). The screenshot combining one mild and one aggressive dark pattern, screenshot J9, obtained more negative scores (M = -1.2, SD = 4.1, Mdn = -1.0, $Mode$ = 0), ranging from -10 to 8. Thus, it attains a lower average rank than the previous two images (M = 2.4, SD = 1.2, Mdn = 2.5, $Mode$ = 1). Screenshot G6, which used the aggressive Visual Interference dark pattern, was rated the least beautiful taking only the absolute scores into account (M = 1.53, SD = 4.9, Mdn = -1.0, $Mode$ = -7), on a range of -10 to 8. However, considering only the relative rankings, its average ranking is similar to that of J9 (M = 2.4, SD = 1.3, Mdn = 2.0, $Mode$ = 1)

The distributions of scores on the complexity scale are depicted in Fig. 3.7. The corresponding stacked bar chart for the complexity ranking is shown in Fig. 3.8 and, again,

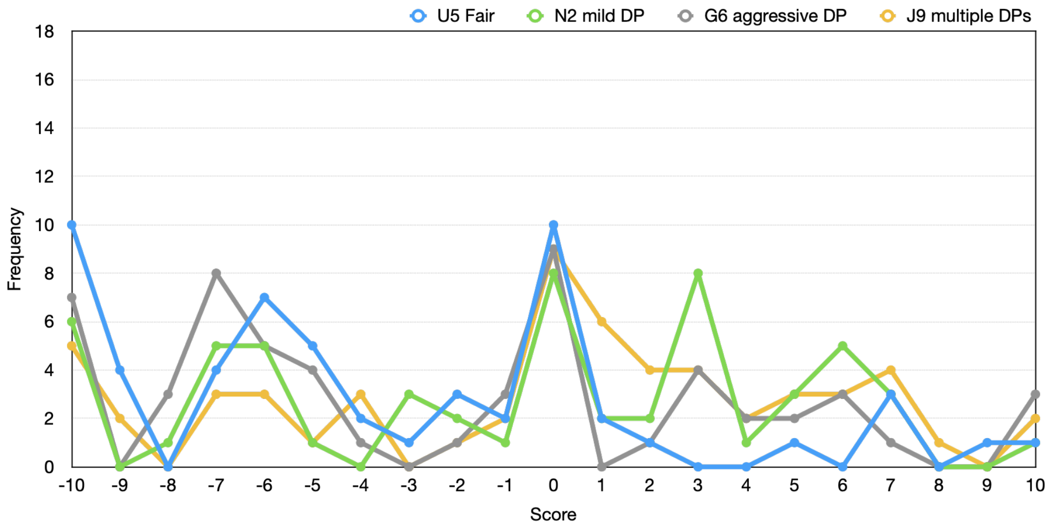

Figure 3.7: The line chart shows the distribution of the complexity scores per design for the semantic differential scales in Part 1 (N = 66). The y-axis depicts the frequency of each score which ranges from -10 (not complex) to 10 (very complex). Each line indicates one of the four designs U5, N2, G6, and J9.

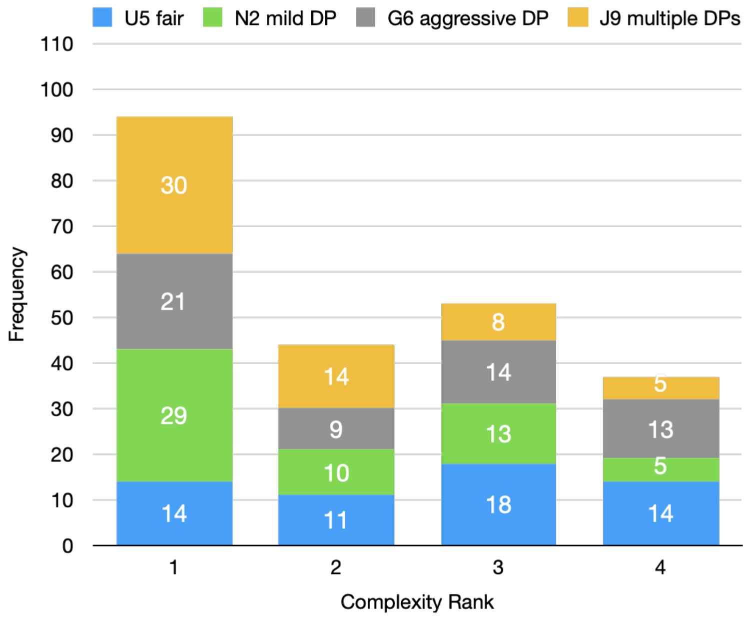

Figure 3.8: The Frequencies of complexity rankings for designs U5, N2, G6, and J9 (N = 66) in a stacked bar chart. It shows the distribution of rankings from 1 (most complex) to 4 (least complex) for each design, represented by different colours.

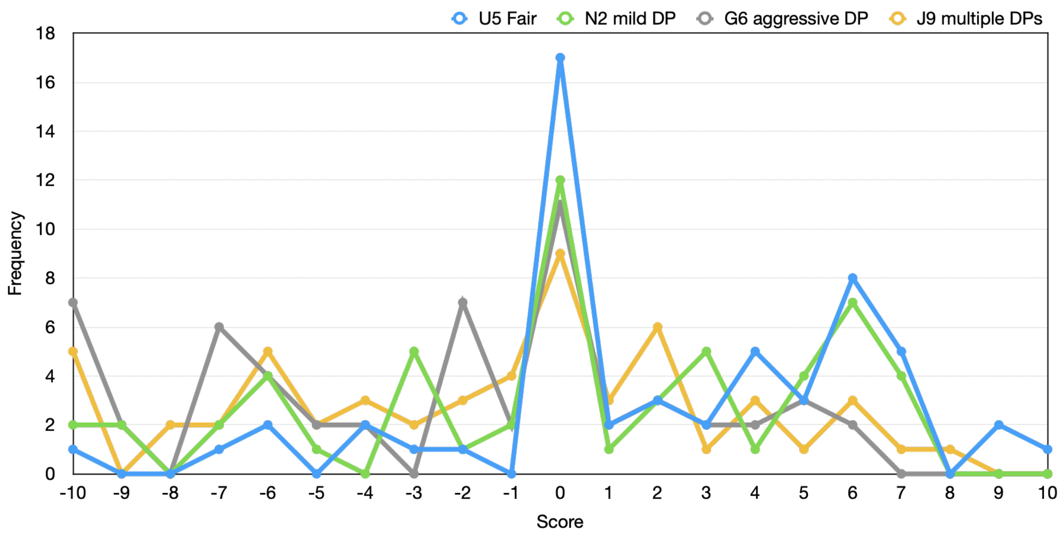

Figure 3.9: The line chart shows the distribution of the trustworthiness scores per design for the semantic differential scales in Part 1 (N = 66). The y-axis depicts the frequency of each score which ranges from -10 (not trustworthy) to 10 (very trustworthy). Each line indicates one of the four designs U5, N2, G6, and J9.

The design with multiple dark patterns J9 was rated the most as complex, the fair design U5 was rated the least

demonstrates that all four ranks were covered for all screenshots. The figures show that J9 with multiple dark patterns obtained the highest but slightly more negative complexity score (*M* = -0.1, *SD* = 5.5, *Mdn* = 0.5, Mode = 0). Likewise the complexity scores of the other three images, its values cover the entire range from -10 to 10. The diagram in Fig. 3.8shows that this screenshot was assigned to the first rank most frequently (*M* = 1.8, *SD* = 1.0, *Mdn* = 1.0, Mode = 1). The second-most complex screenshot was perceived to be N2, using the mild dark pattern (*M* = -0.7, *SD* = 5.5, *Mdn* = 0.0, *Mode* = 0). Its average rank is slightly lower than that of J9 (*M* = 1.9, *SD* = 1.1, *Mdn* = 1.0, Mode = 1). The scores for G6 with the aggressive dark pattern are distributed more strongly on the negative half (*M* = -2.1, *SD* = 5.8, *Mdn* = -2.0, *Mode* = 0). Its average rank suggests that this image was perceived as the second least complex (*M* = 2.3, *SD* = 1.2, *Mdn* = 2.0, Mode = 1). The fair screenshot U5 achieved the lowest complexity ratings (*M* = -3.5, *SD* = 5.4, *Mdn* = -5.0, *Mode* = 0). Its ranking scores are comparatively evenly distributed across all four ranks and reflect the lowest average rank (*M* = 2.6, *SD* = 1.1, *Mdn* = 3.0, Mode = 3).

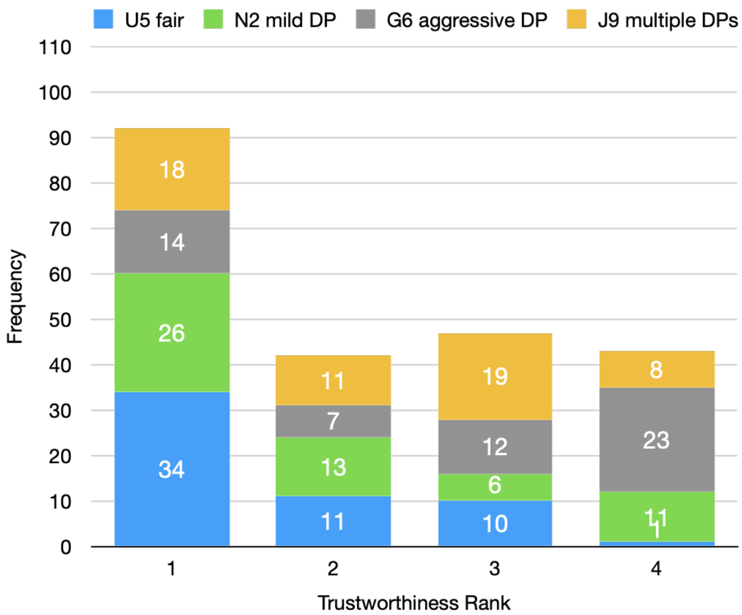

Figure 3.10: The Frequencies of trustworthiness rankings for designs U5, N2, G6, and J9 (N = 66) in a stacked bar chart. It shows the distribution of rankings from 1 (most trustworthy) to 4 (least trustworthy) for each design, represented by different colours.

Fig. 3.9 and Fig. 3.10 present the distribution of the trustworthiness and trust-ranking scores. They, again, show clear differences in the perceptions of the four images, while all four ranks were covered for all screenshots. The highest average trustworthiness scores were achieved by the fair screenshot U5 (*M* = 2.1, *SD* = 4.3, *Mdn* = 2.0, *Mode* = 0), ranging from -10 to 10. On average, it was ranked the most trustworthy (*M* = 1.6, *SD* = 0.9, *Mdn* = 1.0, Mode = 1). N2 with the mild dark pattern obtained, again, the second-highest scores (*M* = 0.3, *SD* = 4.9, *Mdn* = 0.0, *Mode* = 0), within the range of -10 to 7. It achieved a medium average ranking (*M* = 2.0, *SD* = 1.2, *Mdn* = 2.0, Mode = 1). Screenshot J9 with multiple dark patterns obtained negative average trust scores (*M* = -1.4, *SD* = 4.8, *Mdn* = -0.5, *Mode* = 0), on a range of -10 to 8. Its average ranking suggests medium trustworthiness (*M* = 2.3, *SD* = 1.1, *Mdn* = 2.0, Mode = 3). The aggressive dark pattern screenshot, G6, gained the lowest scores (*M* = -2.4, *SD* = 4.8, *Mdn* = -2.0,

The fair design U5 was rated as the most trustworthy, the Visual Interference design G6 was rated the least

Mode = 0), ranging from -10 to 8. On average, it was also ranked as the least trustworthy (*M* = 2.8, *SD* = 1.2, *Mdn* = 3.0, Mode = 4).

There are significant differences between the judgements of beauty, complexity, and trustworthiness for the fair design U5 and the design with multiple dark patterns J9

We used the Friedman rank sum test on all three dimensions beauty, complexity, and trustworthiness. The test suggested a highly significant effect of dark pattern usage on children's beauty perception of a website, $\chi^2(3, n = 66) = 11.58$, $p = .009$. A pairwise post-hoc Dunn test with Bonferroni correction was only significant for dark pattern-free U5 versus aggressive dark pattern G6 ($p = .008$) and dark pattern-free U5 versus multiple dark patterns J9 ($p = .02$). Moreover, the Friedman test revealed very highly significant effects of dark pattern usage on children's complexity perception of a website, $\chi^2(3, n = 66) = 17.825$, $p < .001$. A pairwise post-hoc Dunn test with Bonferroni correction was only significant for dark pattern-free U5 versus mild dark pattern N2 ($p = .04$) and dark pattern-free U5 versus multiple dark patterns J9 ($p = .005$). Lastly, there was a very highly significant effect of dark pattern usage on children's trustworthiness perception of a website, $\chi^2(3, n = 66) = 28.566$, $p < .001$. A pairwise post-hoc Dunn test with Bonferroni correction was very highly significant for dark pattern-free U5 versus aggressive dark pattern G6 ($p < .001$), for dark pattern-free U5 versus multiple dark patterns J9 ($p < .001$), and for mild dark pattern N2 versus aggressive dark pattern G6 ($p < .001$).

3.2.2 Parts 2.1 & 2.2 — Selecting a Manipulative Design

75.8% found the differences, 15.2% did not notice an alternative to watching an ad in G6 in Part 2.1

In Part 2.1, children marked and described the differences between the screenshots U5 and G6. A total of 75.8% of participants correctly identified the differences between the fair and the Visual Interference version: 'In U5, there are several options to reject. In G6, there is only one option, which can be easily missed (as I just did)' (*P13*). For 50.0% of the remaining 24.2%, the justifications suggested that they had not noticed the close button in G6 and therefore did not realise that there was an alternative to viewing advertisements in this version: 'The right one [G6] is worse

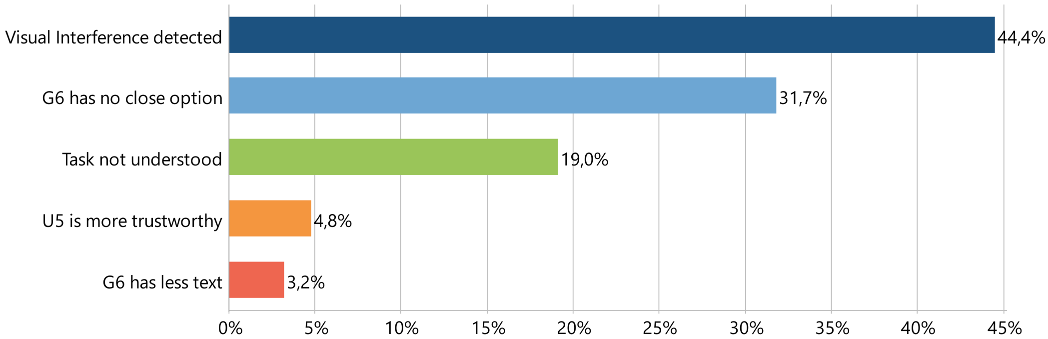

Figure 3.11: The bar chart depicts the frequencies of code categories representing the participants' justifications for their design selection Part 2.2 (N = 65).

because, for example, some people don't like ads and still have to watch an ad. And with the left one [U5] you can decide' (*P22*). Overall, this was the case for 15.2% of all participants. Some participants justified the differences in the trustworthiness of the apps, stating that G6 was 'not trustworthy' (*P58*), while others only mentioned the different labels which served as unique identifiers: 'At the bottom right, there are two different numbers and letters' (*P18*). Other justifications from children who did not correctly identify the differences were missing, incomplete, or implied that the child had not understood the assignment.

In the second task where children could select one design to make people watch an ad, 78.5% selected the Visual Interference design G6, while 21.5% chose the fair design U5. One participant did not answer the question and was, therefore, excluded from the analysis of this task (N = 65). Out of those participants who chose option G6, 84.3% had already correctly identified the differences in the previous part of the task. In contrast, only 50.0% of the participants who chose option U5 had identified the differences, while the other half had not. To understand their motives for choosing one option or the other, we coded and categorised the justifications.

78.5% chose the Visual Interference design

Fig. 3.11 gives an overview of the final set of code categories and occurrences for Part 2.2, representing the justifications for the selections. It shows that the category '*Visual Interference detected*' was the most frequently occurring,

44.4% detected the Visual Interference and its intent

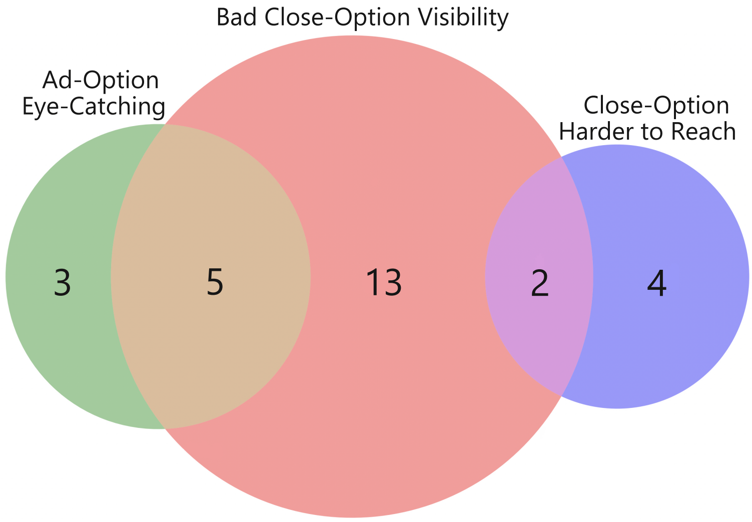

Figure 3.12: Venn diagram illustrating the number of times and overlaps of children mentioning one of the Visual Interference-related arguments *Ad-option eye-catching*, *Bad close-option visibility*, and *Close-option harder to reach* in their justifications for Part 2.2 (N = 65).

Most children justified the selection of G6 by the poor visibility of the close button

with 44.4% of all participants and, respectively, 52.9% of those who selected G6, mentioning any Visual Interference-related design element in G6 and signifying that they were aware of the influencing nature they bear. The category contains three codes capturing different arguments for why the design would fulfil the goal. Fig. 3.12 visualises the intersections between these three contained codes over all participants. The code '*Bad Close-Option Visibility*' covers all justifications mentioning that users might miss the close-option due to its colour, position, size, etc.: 'Perhaps not everyone sees the light cross in G6' (*P05*). As this was mentioned in 74.1% of all justifications related to Visual Interference (30.8% of all participants), it is the most commonly mentioned argument within the category. Several times, it occurred in combination with the second-most frequent Visual Interference code '*Ad-Option Eye Catching*', which refers to the visual dominance of the ad-option, as one 'only gets the one fat option "Watch ad" in the face' (*P13*). Within the category, it was assigned to 29.6% of the responses, and to 12.3% in total. Finally, there was an intersection in two

	G6 selected	U5 selected
Part 2.2		
Visual Interference detected	52,9%	7,1%
Bad close option visibility	39,2%	
Ad-option eye-catching	15,7%	
Close option harder to reach	11,8%	
G6 has less text	3,9%	
G6 has no close option	35,3%	14,3%
U5 is more trustworthy		21,4%
Task not understood	5,9%	64,3%
# N = Documents	51 (78,5%)	14 (21,5%)

Figure 3.13: Cross table providing a hierarchical overview of codes and categories in Part 2.2, showcasing the distribution of codes based on their design selection (N = 65).

cases between '*Bad Close-Option Visibility*' and '*Close-Option Harder to Reach*'. The latter occurred only for 22.2% of Visual Interference-related arguments and for 9.2% in total. It covers all cases in which participants mentioned it was easier to accept watching an ad, despite being fully aware of the existence of the reject option, e.g. because 'there are more options to make people "watch an ad"' (*P61*).

The second-most frequent category of justifications, '*G6 has no close option*' (31.7%), marks all responses suggesting that the child was not aware of the possibility to reject the ad, analogous to Part 2.1: 'With G6 you have no other choice but to watch ads! That is so mean!' (*P16*). Moreover, a considerable amount of children seemed to not have understood the task correctly (19.0%), as their justifications did not match the question: 'If you make all people watch ads, that wouldn't be very good. So I would rather let people choose because otherwise they are getting blackmailed' (*P14*). Out of those participants who selected design U5, it was 64.3%. 4.8% of the children also reasoned their decisions with trustworthiness: 'There is a greater possibility of not watching an ad [in U5], which makes it more trustworthy, so people are more likely to watch the ad' (*P28*). Finally, 3.2% argued that G6 had 'less text, so you decide faster' (*P56*).

Notably many children had not understood the task or did not notice the possibility to reject the ad

A majority of those who chose G6 had detected the manipulative power of the Visual Interference

Fig. 3.13 illustrates how often these codes of justifications appeared in connection with which decision. It shows that the most prominent reason for choosing design G6 with the Visual Interference was that children had correctly suspected the manipulative power of its design elements (52.9%), such as button size and colour. Yet, there was one individual who seemed to have consciously perceived the Visual Interference elements and their potential to influence but still decided on screenshot U5 due to its trustworthiness: 'I chose picture U5 because with G6 you are almost forced to watch the ads and that's why I would reject it' (*P31*). Out of those participants who had detected the Visual Interference, more than 90% had previously correctly identified the differences between the designs. Another frequent justification for choosing G6 was the common misconception that the design did not offer any option to reject the ad (35.3%). Strikingly, the most common code by far among those participants who selected screenshot U5 is '*Task not understood*' (64.3%). We also note that a majority of participants who seemed to have not understood the task also had not found the differences between the screenshots in Part 2.1 (58.3%).

A majority of those who chose U5 had not understood the task

3.2.3 Task 2.3 — Identifying the Goal

66.7% identified that the goal was to make users accept cookies

After coding and categorising the participants' speculations about the goal behind choosing screenshot T3 with the Visual Interference dark pattern, we identified only two different goal categories: Out of all 66 participants, 66.7% of the responses were coded as '*Goal: Accept Cookies*', which includes statements ranging from specific user interaction, e.g. 'Perhaps she didn't want people to see the "reject all" button' (*P04*) to more abstract and profound hypotheses: 'The goal is that the user clicks on "accept all" and thereby shares data' (*P56*). Only one individual provided an alternative goal behind design T3, which we coded as '*Goal: Data Privacy*': 'You should take T3 because all data is protected, but you should not take F7 because the data is not protected' (*P07*). None of the remaining 31.8% specified a goal. Yet, 81% of those participants described the differences between both screenshots but did not accomplish de-

Almost none of the remaining children identified any goal

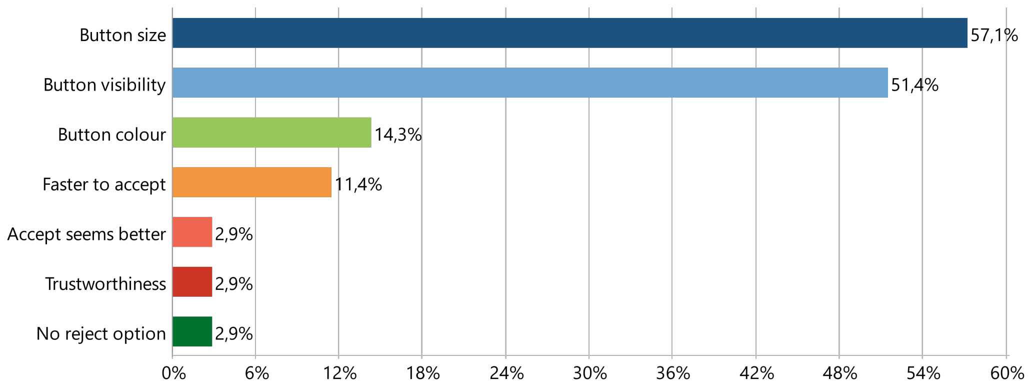

Figure 3.14: The bar chart depicts the frequencies of codes representing the participants' justifications for the goals they assumed in Part 2.3 (N = 34, only participants included who provided a justification).

riving a goal from it: 'In the bottom image [T3], "reject all" is different from the first [F7] and this may lead to a different result, which I don't know exactly' (*P23*).

Since the data privacy-related goal was only mentioned by one child, we coded only the justifications for the '*Accept Cookies*' goal. Fig. 3.14 presents these justification codes, together with their relative frequencies. We found that with 57.1% of all justifications, most children mentioned the different button sizes for accepting and rejecting: 'The "accept all" button is much bigger, so you press it' (*P36*). Almost as often, children mentioned the poor visibility of the reject button (51.4%): 'More people will accept all cookies because the reject button is much less visible! They're so smart!' (*P16*). This argument occurred mostly in combination with other justifications. For instance, in 38.9% of the cases, it was mentioned together with the code '*Button size*', and in 11.1% of the cases with the code '*Faster to accept*'. The latter includes all 11.4% of the cases in which children argued that people would reflexively or out of a hurry press the accept button: 'In T3 you don't see the "reject all" box that well, and people want to see the page as quickly as possible, so they might not see the "reject all" as quickly and press "accept all"' (*P30*). Also, in 14.3% of all justifications, children mentioned the grey colour of the reject button, which makes it less easily visible and might even induce the misinterpretation that the button was in-

'Accept Cookies' goal was mostly justified by different button sizes and visibilities active. This explanation occurred mostly in combination with the 'Button size' code. One child explained that 'because the "reject all" button is so small and almost unreadable, the "accept all" button may seem superior' (*P33*). Another child characterises the Visual Interference screenshot as 'more trustworthy' (*P52*), while *P47* seemingly fell for the grey button colour and did not notice there was an option to reject the cookies: 'T3 doesn't want you to keep your data private, so you can't refuse there'.

33.3% identified the manipulative intent in both parts 2.2 and 2.3, while 24.2% did not link any manipulative intent to the designs

Looking at these results in relation to the results from previous tasks, we discovered that, overall, 33.3% of all participants had both (implicitly) identified the Visual Interference in Part 2.2 and correctly ascertained the manipulative goal in Part 2.3. Looking at it more deeply, 78.6% of those participants who had detected the Visual Interference in Part 2.2, also identified the aforementioned goal and, on the other hand, 50% of those who identified this goal had also detected the Visual Interference in the previous task. A total of 24.2% did not identify either.

3.2.4 Bonus Part — Drawing Manipulations

A majority of drawings included Visual Interference elements

Since we had introduced this part as a voluntary bonus task, we matched the number of participants in the analysis of this task to the 78.8% who had completed it (N = 52). The resulting code categories describing the type of manipulations used in the children's drawings are presented in Fig. 3.15. It shows that 73.1% of all drawings contained at least one dark pattern, while Visual Interference was the most popular, covering 55.8% of all drawings. This category contains different design tricks which are listed in Fig. 3.16, e.g. designing the accept button to be bigger (48.3%) or more colourful (48.3%), or hiding (44.8%) or greying out (24.1%) any alternative option. One example drawing is given in Fig. 3.17 P19, where the reject button is not only clearly smaller and less colourful but it is also placed at an unexpected position as it is not aligned with the accept button.

The second most popular but considerably less common

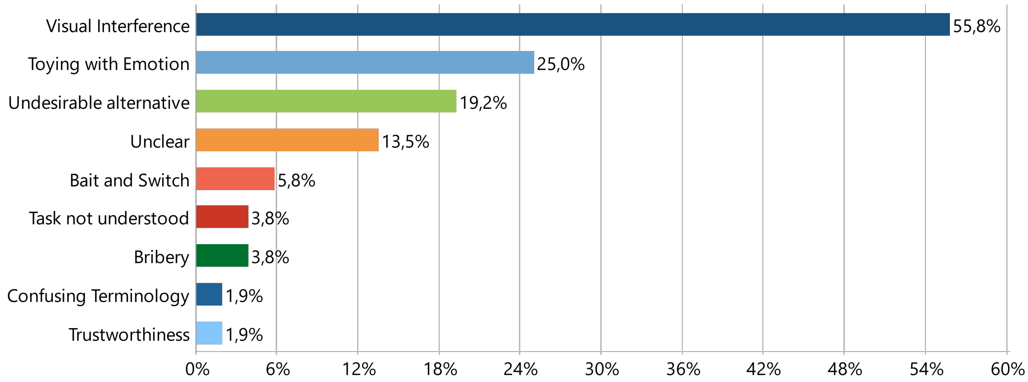

Figure 3.15: The bar chart depicts the frequencies of code categories representing the participants' hand-drawn manipulations in their redesigns in the Bonus Part (N = 52).

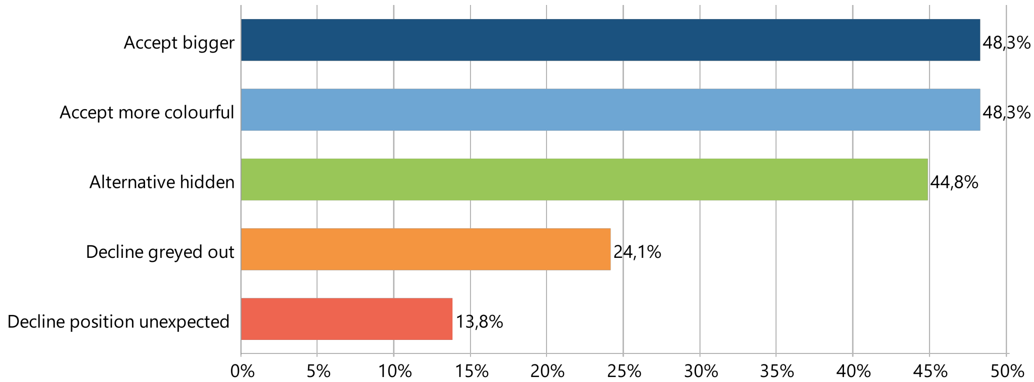

Figure 3.16: The bar chart depicts the frequencies of codes within the Visual Interference category representing the participants' hand-drawn manipulations in their redesigns in the Bonus Part (N = 29).

Using Toying with Emotions or providing undesirable alternatives were also prominent

dark pattern applied in the drawings was Toying with Emotions (25.0%). Out of those cases, 53.9% used influential formulations and 46.2% used smileys and symbols to 'make cookies seem cooler' (*P46*). For instance, Fig. 3.17 P42 uses formulations like 'We use cookies *for a better future*', while in Fig. 3.17 P10, the participant drew happy smiley faces when they mentioned cookies and formulated the text in such a way that made accepting cookies sound like the most favourable option for maximum data security: 'I lied a little bit now so the player thinks everything is safe, even though it's not' (*P10*). Moreover, 23.1% used common colour conventions to communicate right and wrong deci-

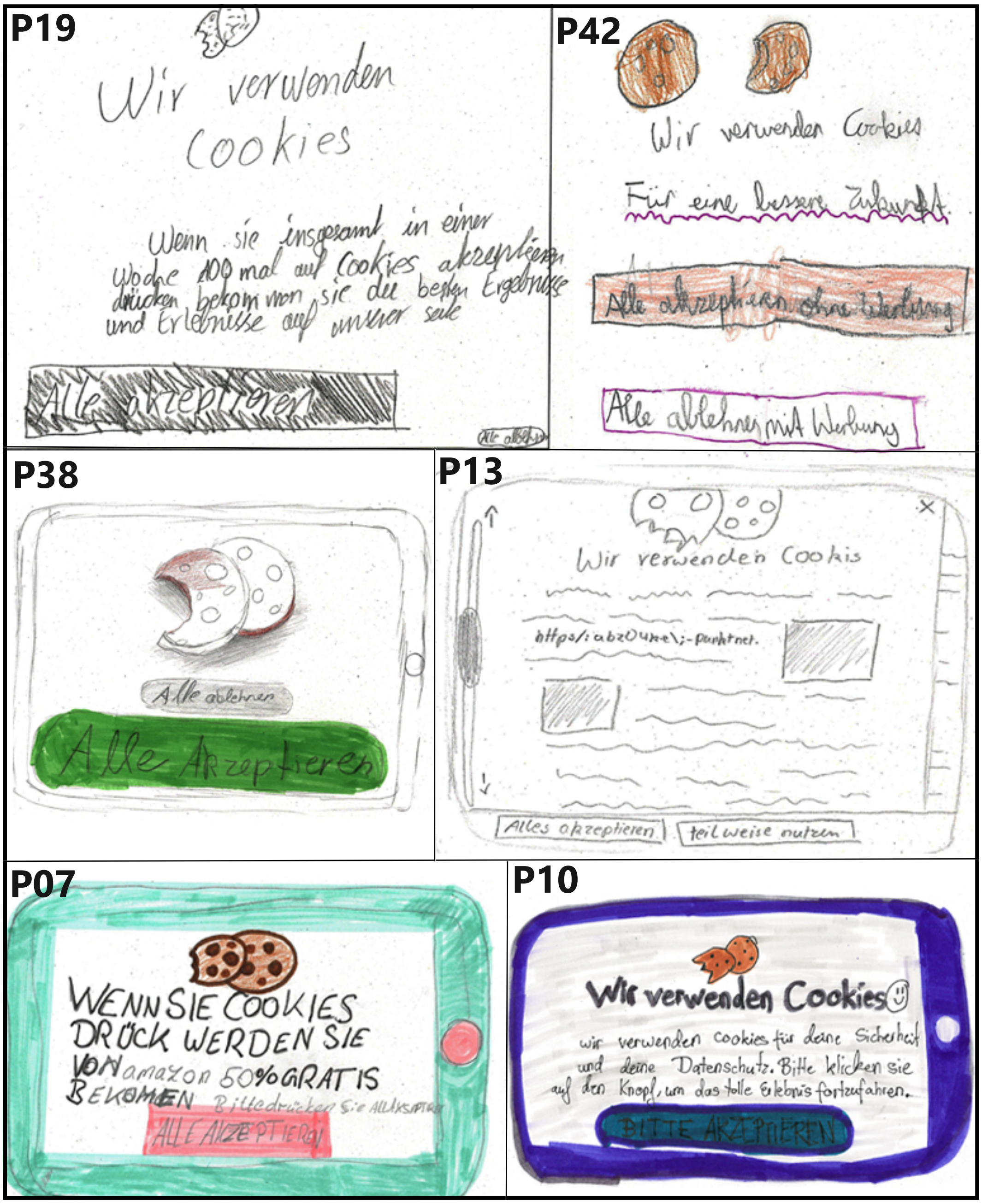

Figure 3.17: A collection of children's drawings from the Bonus Task. *P19, P38,* and *P13* used Visual Interference techniques to make the reject button less visible than the accept option. *P42, P13,* and *P07* try to attract users to accept through a compromise (e.g. Amazon vouchers). *P42* and *P10* use Toying with Emotions with their biasing formulations.

sions, such as using the colour green to signify the 'right' option (see Fig. 3.17 P38).

Another trick that was often used was to provide *'undesirable alternatives'* (19.2%). This includes designs that give no alternative option at all (66.7%, see Fig. 3.17 P10) or those that offer a compromise as an alternative (22.2%), such as providing the option to 'partially' accept (see Fig. 3.17 P13). One child replaced the reject button with a button labelled 'leave app': 'For sure, no one wants to leave the app, so they have no choice but to accept everything' (*P66*), while another child set the condition that one must first watch an ad before one can reject cookies (see Fig. 3.17 P42). Less popular, at 5.8%, was the Bait and Switch dark pattern, where, for example, the reject button carries a small print that invalidates the rejection. Also, 3.8% tried bribery to acquire cookies, such as promising discounts on shopping websites in return for accepting cookies (see Fig. 3.17 P07). Finally, one participant explained that they tried to convey a trustworthy feeling by using smileys to make people accept cookies (*P52*), while another mentioned that one could use confusing terminology to trick the user into accepting (*P13*). A total of 13.5% of the drawings were furthermore coded as '*unclear*' due to missing explanations for non-self-explanatory drawings and 3.8% seemed to not have understood the task correctly.

Often, combinations of tricks were used

While children used a wide variety of tricks and techniques, it is noticeable that most drawings combined multiple of these techniques, as the examples in Fig. 3.17 show. For example, Visual Interference and Toying with Emotions both occurred often in combination with other tricks, e.g. 24.1% of the drawings that contained Visual Interference also involved Toying with Emotions, and 23.1% of Toying with Emotions-drawings also implemented some type of undesirable alternative option. Similarly, there were often combinations of specific tricks within a category, especially within the Visual Interference. For example, out of all Visual Interference occurrences, the codes '*Accept bigger*' and '*Accept more colourful*' were combined in 27.6% of the cases, and 17.2% combined '*Decline greyed out*' with '*Accept bigger*'.

Looking for a connection to previous tasks, we found that

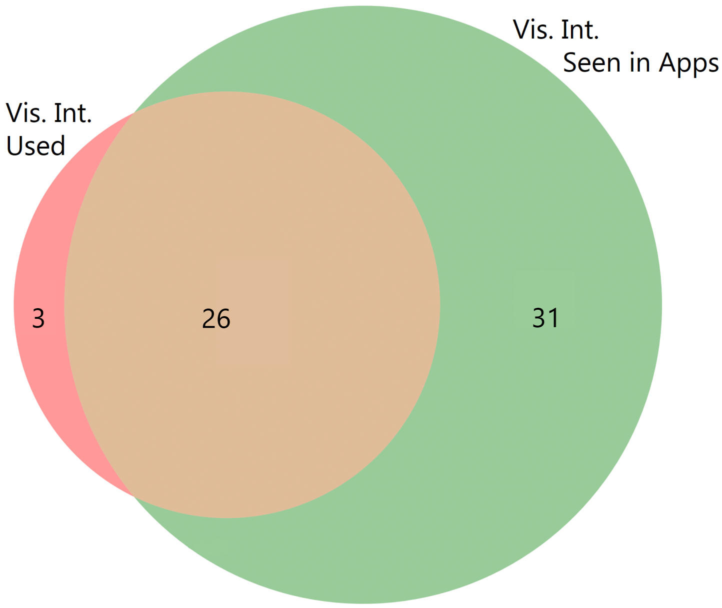

Figure 3.18: Venn diagram illustrating the number of times and overlaps of children using Visual Interference-related manipulations in the Bonus Part and children who have used apps that contain Visual Interference (N = 66).

21.2% detected the manipulative intent in both previous tasks and drew at least one dark pattern themselves

a total of 21.2% of all participants had detected both the Visual Interference in Part 2.2 and the manipulative goal in Part 2.3, and had also applied at least one dark pattern in the bonus part, themselves. Most popularly, 78.6% of these cases included at least one Visual Interference. To gain insight into whether previously frequently used apps and the resulting exposure to dark patterns could be a possible cause for the use of dark patterns in this task, we searched the list of dark patterns known from apps for an overlap with the children who had drawn the corresponding manipulations. As pictured in Fig. 3.18, we found that from all children who used Visual Interference in their drawings, 89.7% had likely seen them before, according to the apps they stated to use frequently. On the other hand, 54.4% did not implement a Visual Interference, although having seen it before in practice. We also found an overlap for Toying with Emotions, albeit notably smaller, as shown in Fig. 3.19. It shows that only 35.7% of the children who used Toying with Emotions in their drawings had likely encountered it

Almost all children who used Visual Interference had likely encountered it before

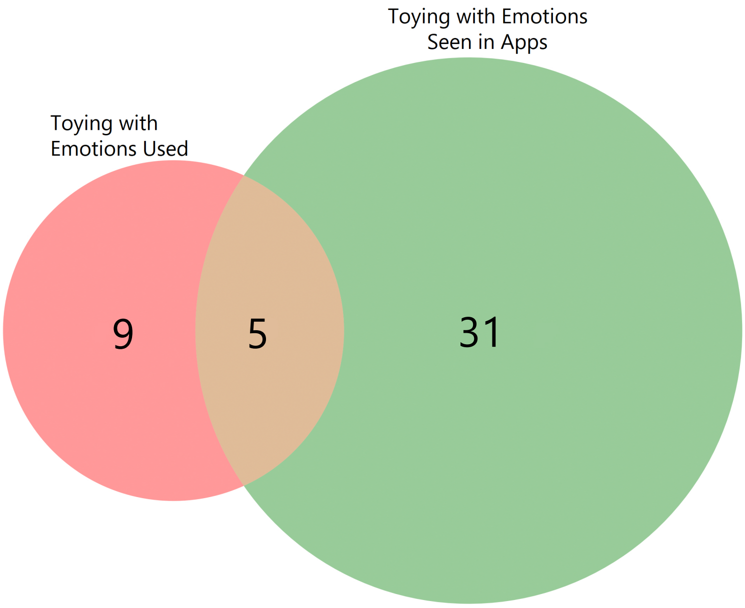

Figure 3.19: Venn diagram illustrating the number of times and overlaps of children using Toying with Emotions-related manipulations in the Bonus Part and children who have used apps that contain Visual Interference (N = 66).

previously in practice, while only 13.9% of all participants who had seen it before transferred their experience to their drawings.

3.2.5 Part 3 — Recognising Manipulations

A total of 54 children responded to the task in which they were asked to mark all manipulations they could find on a screenshot. Hence, we adapted the total participant number in our analysis of Part 3 accordingly. Since we designed the screenshot for this task with its contained dark patterns ourselves, we began our coding procedure with a set of codes derived from the dark patterns Toying with Emotions, Confirmshaming, Trick Question, and Preselection. However, some children noticed and marked other manipulations which we had not intentionally included in our design, resulting in a number of inductively generated

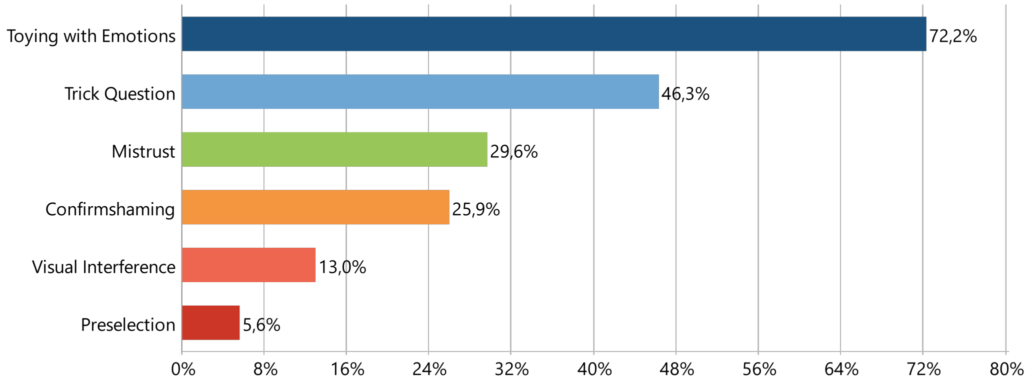

Figure 3.20: The bar chart depicts the frequencies of code categories representing the manipulations the participants identified in Part 3 (N = 54).

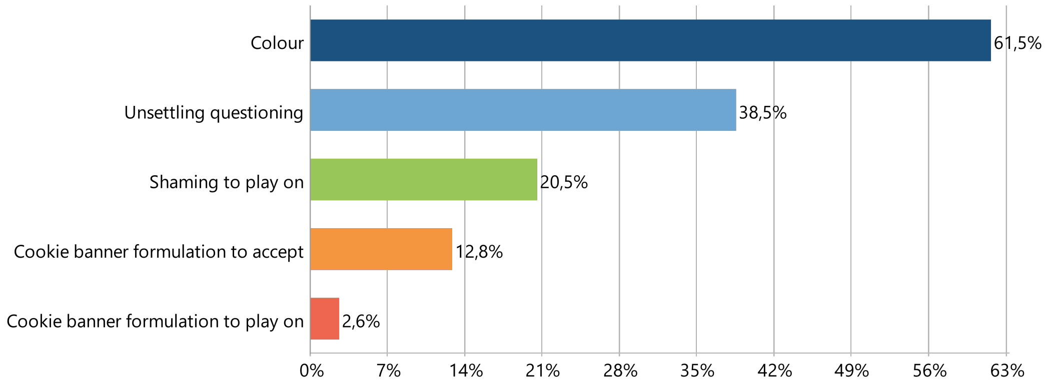

Figure 3.21: The bar chart depicts the frequencies of codes within the Toying with Emotions category representing the manipulations the participants identified in Part 3 (N = 39).

72.2% found Toying with Emotions elements

codes. The categories and their frequencies are presented in Fig. 3.20. It shows that the dark pattern that was most frequently detected was Toying with Emotions (72.2%). However, this category does not only comprise the use of red and green button colours but also other tricks that 'toy' with users' emotions, as Fig. 3.21 lists. Nevertheless, our intentionally designed button colours were marked most often, in 61.5% of the cases: 'Usually, something is red when it is wrong. It's like saying it's wrong to leave the app. And cancel is green, like saying it's right to stay in the app to play' (*P17*). Another 38.5% mentioned that 'sometimes you don't know if you should quit the game' (*P06*) when asked again if you are sure, so they marked the prompt in the pop-up window as being manipulative. Moreover, in 20.5% of the responses within this category, the formulation 'I play too bad' from the second check box was marked as manipulation to make users play on 'because then you are motivated to play more and also better' (*P12*). Finally, some children suggested that the cookie banner formulation 'for the best experience on our site' was intentionally formulated in a way to 'make it look like something cool is going to happen when you press ok' (*P09*, 12.8%), or 'so people think the game is somehow cool and keep playing' (*P36*, 2.6%).

Besides Toying with Emotions, the second-most frequently identified dark pattern was the Trick Question in the cookie banner formulation, which was mentioned by 46.3% of the responses. Furthermore, 29.6% expressed a general mistrust towards specific design elements, without linking them to concrete manipulations. For example, one child marked the 'are you sure [...]?' question, annotating that it 'sounds like it's dangerous' (*P33*). Other sources of mistrust were, in 50.0% of the cases, the mere existence of the cookie banner, interlinked with misunderstandings due to them not having noticed the Trick Question: 'It's there to annoy the player a bit and to make the player think: "So what, as long as the purple box is gone" and press ok' (*P10*). Also, 25.0% expressed mistrust as they were concerned over data privacy issues regarding the option to share points with friends: 'If you share points, the data can be easily intercepted' (*P34*). Within this share points option formulation, a total of 25.9% of all responses revolved around the Confirmshaming we placed there: 'You would rather share it with your friends than admit that you play badly' (*P42*). Also, 13.0% mentioned the different button sizes, claiming that the ok button was 'much smaller and not as eye-catching', which we categorised as Visual Interference. The final dark pattern we included in our design was Preselection, which was detected by the fewest participants with 5.6%.

The second-most frequently detected dark pattern was the Trick Question

Many children expressed general mistrust towards the app

Confirmshaming was identified by 25.9%, while Preselection was least frequently found

We further clustered the responses to investigate common combinations of the four dark patterns from the questionnaire found by the participants. The clusters showed that, out of all 66 participants, 25.8% had not found any dark

25.8% did not find any dark pattern and nobody found all four

Trick Question and Toying with Emotions were often found together

pattern, while nobody had found all four. 28.8% identified solely the Toying with Emotions pattern. Confirmshaming was always detected in combination with the Trick Question or Toying with Emotions, but never alone or together with Preselection. Another 10.6% found the most frequent combination of two: Trick Question and Toying with Emotions. 13.6% identified Confirmshaming, Toying with Emotions, and Trick Question, at once. These proportions considerably contract when considering only those participants that had already successfully detected the Visual Interference from Part 2.2: out of those 28 children, only 17.9% did not manage to find one of the dark patterns in this task and no combination of two patterns was found by more than three children. However, the percentage of children who had only detected the Toying with Emotions pattern had slightly increased to 35.7%, just as the combination of the three patterns Confirmshaming, Toying with Emotions, and Trick Question, which increased to 14.3%.

Chapter 4

Discussion

The objective of our study was to enhance our comprehension of children's mental models concerning dark patterns. In particular, we wanted to investigate whether children can recognise the deceptive goals behind dark patterns and how their prevalence in our everyday lives plays a role in this. Our results show that while many children have some basic understanding of the manipulative power and intent behind certain design elements on websites and apps, other children have not been able to make this connection. Moreover, the level of understanding may vary depending on the dark pattern type. However, we could not confirm a connection with the dark patterns that the children had presumably seen before.

In the following, we discuss and interpret our findings from different perspectives, compare them with existing knowledge from previous work and look at their implications for possible ways to protect children from dark patterns in the future. Finally, we review the limitations of our study and their impact on the interpretation of the results of this work.

4.1 Spontaneous Judgements of Websites Using Dark Patterns

We first wanted to assess children's spontaneous judgements of websites using different types and degrees of dark patterns, ranging from no dark pattern to combining both mild and aggressive ones in one screenshot. Our results reveal that the dark pattern-free screenshot U5 was overall judged most favourably, i.e. on average, it was perceived as the most beautiful and trustworthy, and the least complex design. On the other hand, either design G6, which used solely the aggressive Visual Interference dark pattern, or design J9, which combined this with a mild Confirmshaming dark pattern, were always judged least favourably. It is interesting to note that U5 and J9 were the only pair to show significant differences in all three aspects, as these designs are exactly the two opposites with respect to the degrees of dark patterns used. This finding is in line with the observations by Conti and Sobiesk [2010] from their study with adult participants: they found that users would develop overall negative attitudes towards websites when they realised they entailed manipulative intents. This could be the first indication that the children in our study did have a spontaneously manipulative impression of the designs that used aggressive dark patterns. Interestingly, this impression hardly seems to exist in designs with Confirmshaming patterns, which suggests that different dark pattern types are spontaneously and probably unconsciously perceived as differently manipulative. This discovery is also confirmed in other questionnaire parts and will be taken up again later in our discussion.

Children judging designs with dark patterns more negatively is in line with findings from related work with adults

Looking at the three attributes of beauty, complexity, and trustworthiness separately, we can make further observations. The average beauty scores indicate the following order from most to least beautiful: U5, N2, J9, G6 — which is in accordance with Mocka [2022]'s finding that (adult) users perceived apps with dark patterns as less appealing. However, looking at the rankings, we found that all four screenshots were most often ranked highest (i.e., most beautiful). While this may sound counterintuitive, the reason was that we allowed several screenshots to be assigned

Designs containing aggressive dark patterns were on average rated as less beautiful, which is in line with findings about adults from related work the same position on the scale. This opportunity was used comparatively often by our participants, resulting in several or even all screenshots being assigned the same rank, which, in this case, was 1 by default. This may be due to children not having understood the task, or simply because they actually judged all designs equally. While the rankings for N2 were relatively equally distributed among the remaining ranks 2 to 4, design G6 was ranked most controversially, as it was most frequently assigned rank 1 and almost as often the final rank, rank 4. This could, again, be explained by the findings from Conti and Sobiesk [2010] and Mocka [2022]: while some children may have already noticed the deceitful intent behind the design and rated it poorly, in accordance with previous findings from the literature, some others may not have realised the Visual Interference and its manipulative power, therefore not ranking it as badly. The fact that a considerable proportion of children was not able to identify and correctly interpret the Visual Interference was also shown in later tasks.

Children who ranked the Visual Interference design high may not have realised its manipulative intent

While for the beauty scale, all designs which contained a Visual Interference scored the least, the complexity scores reveal a different order with a rather negative judgement towards designs using Confirmshaming. Going from most to least complex, the order was J9, N2, G6, and U5. Notably, all means were negative, i.e. all designs were rated as tendentially not complex. Particularly U5 and G6 received predominantly negative scores and were thus explicitly perceived as less complex. This could be explained, as before, by children not having noticed the manipulative power behind the Visual Interference in G6, wherefore it was judged as similarly non-complex as other dark pattern-free designs, such as U5. This, again, is in line with the literature stating that users evaluated apps with dark patterns as more complex to use [Mocka, 2022], whereas no negative impact on the judgement of an app was observed when users were not aware of any manipulative intent [Conti and Sobiesk, 2010]. On the other hand, those designs that contained Confirmshaming, J9 and N2, were most frequently and almost equally often assigned rank 1, i.e. most complex. Apart from the dark pattern-related reasons suggested by the literature, an alternative explanation for this finding could be that Confirmshaming usually

Confirmshaming designs were perceived as more complex, probably due to visual clutter, as suggested in the literature comes along with longer sentences. This may lead to visual clutter which, again, is a complexity determinant and may, thus, result in higher complexity ratings for these designs [Miniukovich and De Angeli, 2014]. Moreover, it is worth highlighting that there were significant differences between the judgements of U5 and G6 for both beauty and trustworthiness, but not for complexity. This suggests that the Visual Interference may not impact the perceived complexity as much as Confirmshaming does.

Visual Interference was judged as not trustworthy, in accordance with prior findings about adults, whereas Confirmshaming did not negatively impact trust perceptions

Finally, the trustworthiness scale attained the same order of average scores as the beauty scale, being U5, N2, J9, and G6, from most to least trustworthy. Notably, both U5 and N2 had positive means and were most frequently assigned the first rank, while J9 and G6 were rated rather negatively, i.e. as non-trustworthy. This corresponds to the findings from related literature about adults, indicating that they tended to have less trust when encountering dark patterns [Voigt et al., 2021]. Alternatively, one reason why the two Visual Interference designs obtained poor trustworthiness scores may be that, as later tasks reveal, a considerable number of children had not noted the option to reject watching an ad and had, hence, misinterpreted the design as not giving the user a choice. Several children later stated that, because of its supposedly forced action, they perceived design G6 as not trustworthy. Also, there were very highly significant differences between the trustworthiness judgements of N2 and G6, implying that Confirmshaming does not seem to negatively affect children's trust perceptions, in contrast to prior findings from the literature [Ekroth and Sandqvist, 2020]. Still, it supports Luguri and Strahilevitz [2021]'s findings that aggressive dark patterns generally received more powerful backlash from users than mild dark patterns.

To conclude, our findings about children's judgements of designs using dark patterns were roughly in accordance with findings about adults from related work. However, we gained new initial insights into how different dark pattern types differently affect children's judgements: while Visual Interference designs were rated significantly more negatively regarding their beauty and trustworthiness, they were not perceived as complex, as opposed to the Confirmshaming design. Notably, we also pointed out several potential alternative dark pattern-independent explanations for our findings.

4.2 Children's Understanding of the Intent Behind Dark Patterns

One of the questions to be answered when exploring children's mental models of dark patterns is whether they can recognise the manipulative intention or even the danger behind certain design elements. To explore their understanding of this relationship, we investigated in Parts 2.1–2.3 whether children could select the manipulative design that matches a goal and, conversely, identify a goal that matches a design. The results of our study indicate that, overall, every third child did understand the manipulative power and deceitful intent behind designs using Visual Interference dark patterns, while every fourth was not able to make this connection, in none of the tasks. Moreover, almost none alleged any manipulative intents in the dark pattern-free design.

Taking a closer look at the results for Parts 2.1–2.2 we reported earlier, we found that almost half of the participants had justified their responses in a way that we can assume with relatively high confidence that they have indeed realised that Visual Interference can influence user decisions and how. However, this cannot be generalised unconditionally, as the children in this task were actively seeking designs to fulfil a given (manipulative) goal. It remains to be researched how these proportions would shift if children were only unconsciously exposed to such tricks. Moreover, while already roughly 80% of all participants had chosen screenshot G6 with the Visual Interference to match the goal, we noted that, interestingly, almost 86% of the remaining children who chose dark pattern-free U5 had responded in a way that we assume they had not understood the task. Inversely, also 75% of the children who did not seem to have understood the task chose option U5. On the one hand, this observation could be based on order effects,

Children's cognition of dark patterns, especially Visual Interference, needs to be researched when unconsciously exposed to them i.e. children simply selecting the first option, or on personal preferences — independent of the question regarding the goal — since we have already seen that U5 was generally preferred over G6. While we can only speculate about the reasons, it is interesting to point out that if we exclude the results of those participants who did not understand the task, almost 90% of the participants would have chosen G6 with meaningful justifications. Subsequently, only two children chose design U5 while giving reasonable arguments for why this design would make users watch an ad. Both justified their selection with U5 being more trustworthy wherefore they would rather watch an ad when exposed with this design than with G6. Hence, neither these two participants nor any other child in the study had erroneously inferred any manipulative intent behind design U5.

Only two children reasonably argued for the manipulative effectiveness of U5

Furthermore, we detected throughout the entire questionnaire that almost every third participant had not seen the light close button in design G6 and, thus, missed the option to reject watching an ad. While we have not tested children interacting with the designs in our study, we can assume that these children would have likely fallen for the Visual Interference if they encountered this situation in real life, leading them to (involuntarily) watch the ad. Although we have no comparative measurement for dark pattern-free designs, we can still expect a relationship between exposure to Visual Interference and the child's decision to watch an ad or not. This would be in accordance with Luguri and Strahilevitz [2021]'s findings that aggressive dark patterns were considerably more effective than dark pattern-free user interfaces in terms of preference inconsistency. Also, Lupiáñez-Villanueva et al. [2022] found that not all users were equally susceptible to malicious designs, supporting our discoveries on children.

One-thirds of the children would have likely fallen for Visual Interference in an interactive setting

Our previous findings were further supported by the results from Part 2.3 of the questionnaire. The data revealed that two-thirds of the participants were able to correctly identify the malicious intent behind the Visual Interference in the cookie consent design, being to make users accept all cookies. More than that, apart from one child, nobody had identified an alternative goal. While this could again suggest that children have an intuition for how aesthetic manipulations in designs may affect user decisions, conclusions should be made cautiously. It is important to note that, until this task, the participants had already seen and actively reflected on several manipulative designs within the questionnaire, particularly Visual Interference. Although each child had worked on the previous tasks individually without exchanging their thoughts, this might have shaped their mental models accordingly, threatening the generalisability of our results. Additionally, most children were already somewhat familiar with the concept of cookies, or at least with the fact that one should be careful with unquestionably accepting them. Knowing this, the cookie banner scenario may have immediately alerted the children and steered them to think in this direction. Nevertheless, their answers were well-justified, mentioning typical Visual Interference elements, such as differing button colours, sizes, etc., wherefore we can rule out mere guessing and assume a certain level of understanding.

Seemingly broad understanding of Visual Interference may also be due to the children being familiar with cookie banners

Overall, every third child showed a general understanding of the connection between interface elements and manipulative power in both tasks 2.2 and 2.3 dedicated to this question. On the other hand, every fourth child was unable to link the two, making them potentially susceptible to dark patterns or, at least, to Visual Interference tricks. Investigating this phenomenon with different dark patterns, apart from Visual Interference, and other scenarios may reveal further insights and help comprehend which dark pattern characteristics impact children's understanding more or less. This may introduce new possible starting points for effective dark pattern interventions targeted at children.

4.3 Children's Spontaneous Use of Dark Patterns in Redesigning UIs

To further explore the relationship between UI design and manipulative goals, we also involved the children in a voluntary drawing task to redesign UIs in such a way that they would achieve a given manipulative goal. We were particularly interested in whether children would already autonomously incorporate common dark patterns which they had likely seen before due to their prevalence on the internet and apps. We found that three-quarters of the drawings contained at least one dark pattern. More than half of all drawings included Visual Interference elements, while we can assume that almost all children who applied it had encountered this pattern before.

What is conspicuous is that, out of all tricks, Visual Interference was used most frequently. While this may indicate a particularly deep intuition of the children about how this dark pattern is used to manipulate users, we must consider that they were already confronted with Visual Interference several times in previous tasks, which might have biased them when redesigning the screenshot. It would be interesting to repeat the same task in a less-biased environment to inspect whether similar trends would be detected. On the other hand, as mentioned before, the children were already mainly familiar with the concept of cookie banners and have likely encountered common tricks to make users accept cookies which often include Visual Interference, e.g. hardly visible reject buttons. Thus, one possible explanation for the popularity of Visual Interference in this task could be that this design pattern became so ingrained in children's minds that they applied it by themselves, probably even unconsciously. However, we could not draw any further conclusions on the role of the prevalence of dark patterns, based on the apps they frequently used. The reason for this is that almost all children had indicated using apps that used Visual Interference, yielding major intersections with both children who had and had not used this dark pattern. Conversely, more than half the children had likely seen Toying with Emotions tricks before, yet, only very few of them had applied corresponding tricks. Researching this particular question about the impact of the prevalence in a dedicated study might yield novel insights for future research.

Visual Interference was drawn frequently which may be due to its prevalence in real-life cookie consent forms

Moreover, we noticed that most children combined several manipulation techniques in one UI, most commonly Visual Interference and Toying with Emotions. This also has a strong resemblance to the realistic use of dark patterns, as dark patterns are rarely found in isolation in practice [Lupiáñez-Villanueva et al., 2022]. It remains unclear whether there is a connection between this experience of the child and the resulting drawings. Finally, while three-quarters of the drawings contained at least one dark pattern, only every fifth child had detected the manipulative intent in Parts 2.1–2.3 plus applied at least one dark pattern themselves in the Bonus Part.

Several dark patterns were often drawn in combination which resembles real-life scenarios

Overall, this task demonstrated that most children were able to autonomously devise and employ manipulative tricks, which requires a deeper understanding of how users can be influenced to make certain decisions through targeted UI design characteristics such as colour, layout and wording. However, it is important to consider that previous tasks may have introduced biases in favour of Visual Interference. With the methods we used for this task, we were unable to draw any clear conclusions from the redesigns about the potential influence of the prevalence of dark patterns.

4.4 Recognition of Dark Pattern Characteristics

Finally, we wanted to research children's mental models of the term 'online manipulation' and find out which dark patterns, or which of their characteristics, they would recognise particularly well and intuitively, and which less. While three-quarters of the participants identified Toying with Emotions elements as being manipulative, one-quarter of them could not find any manipulations. In some cases, UI elements were declared as manipulative which we had not intended as such in the design of the screenshot.

While earlier in our discussion we had often suspected conspicuous trends in the data to be influenced by designs shown in previous tasks, this suspicion is, at least for this task, refuted: although neither Toying with Emotions nor Trick Questions had been part of the study until now, these two dark patterns were by far the most commonly de-

Most frequently detected dark patterns could not be traced back to experience from previous tasks

tected manipulations in this task. Interestingly, we have already observed throughout the entire questionnaire that also other kinds of aesthetic manipulations, i.e. manipulations in the form of colours, different button sizes and layouts, seem to be particularly intuitively understandable to the children. They were notably prominently the topic of discussion in the results of the previous tasks, including the recognition of manipulative goals and the autonomous design of manipulations. We assume the ability to detect aesthetic manipulations, including Visual Interference and Toying with Emotions, to come with the cognitive development of children 10–11 years old. This would explain why the use of red and green colours in the Toying with Emotions dark patterns were so often recognised in this task. Additionally, half of the participants recognised linguistic tricks, including Trick Question and Confirmshaming. Although not as prominent as the aesthetic manipulations, these seem to be rather easily comprehensible, as well. However, as discussed earlier, we assume that this finding is particularly susceptible to being influenced by the fact that the children were explicitly searching for manipulations. Hence, we emphasise once again the importance of a corresponding study in which children subconsciously interact with such dark patterns. Nevertheless, the observation that dark patterns differ in the level of transparency and effectiveness is in lineage with the related literature [Luguri and Strahilevitz, 2021].

Half the children recognising linguistic manipulations when explicitly searching for them cannot be unconditionally generalised to real-life scenarios

Besides the dark patterns we intentionally designed for this task, other manipulative designs were several times identified, for example, the fact that users were asked whether they were sure about leaving the app. Although this cannot be directly classified as a dark pattern, we consider this finding relevant because the identified design can certainly be ascribed to a manipulative goal, which the children recognised. Especially since the decision as to whether or not a design is a dark pattern can be subject to interpretation [Di Geronimo et al., 2020], this finding contributes to answering our research question. Similarly, when exchanging experiences with the classes at the end of the study, we noted that, although children did express personal experiences with dark patterns, some of their anecdotes much rather described criminal internet fraud than

The distinction between general online manipulations and dark patterns remains unclear to most participants but it does not pose a limitation to our work actual dark patterns. While this proves an erroneous understanding of the exact definition of dark patterns, we do not consider this to be a major limitation of our previous results. Furthermore, a considerable number of children were generally suspicious of the mere existence of the cookie banner, without being aware of the Trick Question it entailed. This was likely due to the fact that the children had already talked about cookies and data privacy in school and were, thus, generally alarmed. It has already been suggested by Zhao et al. [2019] that most children had a basic level of understanding of online privacy and online security. Although we did not intend to investigate this further, we think this is a promising finding that demonstrates the power of dedicated children's education. However, in the discussion round after the study, children explained that they had encountered dark patterns regularly and were even aware of them, but often deliberately deferred to them out of convenience or impatience. While this self-report may not be highly reliable, it aligns with related findings about adults, which revealed that, although being aware of dark patterns, they often were not able to resist them [Bongard-Blanchy et al., 2021]. This assumption with respect to children is yet to be examined in future work.

Prior knowledge might have impacted the results and alerted the users to be careful in certain scenarios on the internet

4.5 Implications for Safeguarding Children from Dark Patterns

Gaining a deeper understanding of children's mental models of dark patterns allows us to draw some implications for this field of research. In the following, we will discuss these with a particular focus on approaches that aim to safeguard children from dark patterns. These include ethical child-friendly design, policies and regulations, and digital literacy education (see Section 2.1.4 "Fighting Dark Patterns").

First, we noticed that mental models can vary greatly among children and can range from one extreme to another. Although in our study we did not assess whether children would fall for dark patterns or not, we observed that many were not able to detect manipulations even when they were

More laws are needed to prohibit the use of dark patterns on apps and websites targeted at children

explicitly searching for them. This leads us to assume that there is a considerable number of children who are particularly susceptible and at risk of dark patterns. Consequently, when thinking about possible approaches to counteracting them, we must especially consider the needs of the most vulnerable groups. One step would be to adapt Jarovsky [2022] and Weinzierl [2020]'s suggestions and extend policies and regulations to prohibit the use of dark patterns on apps and websites targeted at children. Instead, designs should be ethical, transparent, and age-appropriate.

Countermeasures against dark patterns need to be customised depending on the type of pattern

Additionally, we observed different mental models and vulnerabilities depending on the type of dark pattern or specific characteristics of manipulative designs. In the fight against dark patterns for child protection, it is therefore essential to consider different pattern types individually and to customise countermeasures accordingly, as Bongard-Blanchy et al. [2021] also proposed. Especially, one should focus on identifying and ruling out those dark patterns that children are highly prone to fall for. For example, we observed that aesthetic manipulations were overall intuitively understood and, when actively searching, easily detected. Such patterns could therefore probably be efficiently counteracted through awareness-raising, as suggested by Di Geronimo et al. [2020]. Other pattern types like Preselection and Confirmshaming, in contrast, were not so obvious to find for the children when searching for manipulations. Hence, these patterns require to be particularly regulated by law in child-directed UIs. Following up on this, it might make sense to tackle the problem at its roots and encourage the implementation of ethical design. As explained earlier in our work, design students should be made aware of the potential consequences of using malicious design techniques and education of ethical design principles should be included in their curriculum [Gray et al., 2018].

The implementation of ethical designs to reduce dark pattern encounters needs to be encouraged

Include dark pattern education in the school curriculum to make children aware of risks and learn how to resist them

Moreover, in Section 2.1.3 "Impact on the User & Awareness" we have already discussed the impact of awareness on adults recognising dark patterns. In our study, we obtained evidence that through education and greater awareness of dark patterns, children were also more confident in identifying them and put less trust in the website incorporating them. However, we also learned that they would sometimes recognise online manipulations but not make an effort to resist them, similar to what is known about adults [Bongard-Blanchy et al., 2021]. The implication is that children, as well as parents and teachers, need to be educated about dark patterns and malicious designs on apps and websites. The topic should be taken up by schools and taught as part of the digital literacy curriculum. It is also important that children can practice how to handle dark patterns correctly under parental guidance [Kumar et al., 2017], empowering them to recognise and resist manipulative design tactics in future encounters. Particularly, children need to be educated about why it is important not to fall for such manipulations and what the possible consequences could be [Bongard-Blanchy et al., 2021]. We believe that this might impact the way children will deal with dark patterns in future interactions.

4.6 Limitations

When interpreting our findings and deducing implications for the field of research, it is important to be aware of the limitations and complexities of our research. While we have already mentioned several challenges faced during the design of the study, we will discuss the main limitations of our work below.

Most importantly, it should be noted that the combination of mental models research and studies with children is already challenging in itself. In Sections 3.1.1 "Studies with Children" and 3.1.1 "Existing Mental Model Elicitation Techniques", we extensively discussed the advantages and disadvantages of different mental model elicitation techniques, as well as threats to validity and reliability when conducting a study with children. Despite considering these in our study design, there remain shortcomings that we need to be aware of. On the one hand, mental model researchers stress the near impossibility of eliciting perfect, distinct, and complete mental models [Norman, 1983], which principally puts the validity of our results into question. Additionally, there are a number of

Eliciting complete and valid mental models is practically impossible limitations every researcher has to consider when studying children: inconsistencies in children's responses depending on the method and task formulation have been demonstrated [Panagiotaki et al., 2006], as well as a particular susceptibility of children to demand characteristics and pressure [Punch, 2002]. We tried to reduce the influence of these limitations by, for example, triangulating different types of tasks and having responses explained textually, putting the entire questionnaire in the context of a game to reduce pressure, and creating a familiar environment through the setup in the classroom and the presence of the class teacher. Nevertheless, a certain distortion of the results cannot be ruled out.

Children are especially prone to biases

Moreover, the types of tasks we used in the questionnaire might have introduced a bias in the results, since it made the children consciously look for manipulative design elements. This may not accurately reflect their behaviour in a real-life situation and might have, erroneously, suggested a too high detection and accuracy rate. Other factors that may have contributed to this are possibly overly leading questions that helped the children think in the 'right' direction. However, as described in Section 3.1.2 "Questionnaire Creation", we considered this to be the most appropriate method, in trade-off with not overwhelming the children with too abstract tasks. Also, the fact that recurring dark patterns were used throughout the questionnaire (e.g. Visual Interference was present in almost all tasks) could have triggered a learning effect, which might have particularly increased the Visual Interference detection rate. However, we deliberately decided against methods for counterbalancing order effect as the tasks were intended to build on each other and were planned to be in the same order for all children in the interest of a non-stressful atmosphere during the study. For greater focus and deeper insight, we also decided against examining too many different dark pattern types at once and, thus, explored mostly Visual Interference for Parts 2.1–2.3. Yet, we had not intended to measure children's interactions with dark patterns but rather gain initial insights about whether children generally comprehend the potential of using specific design elements with manipulative motives. Therefore, we believe that this limitation poses only a little risk to the validity and importance of our

High detection rates probably due to leading questions, learning effects, and explicit search for manipulations results.

Finally, it is worth mentioning that our results can only conditionally be generalised to the population of 10–11-year-olds since we recruited all of our 66 participants from the same school. As we discovered, the curriculum at the school included basic lessons about internet safety and related topics, so the children had already been educated or at least warned about the dangers of the internet. Although the topic of dark patterns or manipulative designs on the internet was not covered, one may question whether our participants should be considered a representative sample for all German children in this age group in terms of their prior knowledge. Conducting the study with participants from different backgrounds might yield novel and more generalisable results.

Our sample was educated about the basics of internet safety which might reduce their representability for the population

In the course of our work, we attempted to always stay aware of these limitations and to be cautious about generalising our findings to the whole population. nevertheless, this allowed us to gain valuable insights into the field of research and opens up various starting points for future research, which will be elucidated later.

Chapter 5

Summary and Future Work

In the following, we provide a summary and an overview of the contributions of this work. We conclude the thesis by proposing future work that builds up on our findings.

5.1 Summary and Contributions

To extract children's mental models of dark patterns, we collected questionnaire data from 66 fifth graders

In this thesis, our primary objective was to gain deeper insights into children's mental models of dark patterns. The focus of our work was twofold: first, we wanted to investigate children's ability to recognise the deceptive intents behind dark patterns. Second, we wanted to explore how the prevalence of such malicious designs is reflected in children's mental models. For this, we conducted a study with three fifth-grade classes from a local grammar school, involving a total of 66 participants aged 10–11 years old. The children completed a multi-part questionnaire that we designed to extract these models from different perspectives and in a playful manner.

The results reveal that mental models vary among children, but also depending on the characteristics of the different types of dark patterns they were exposed to in the screen-

Some children understood manipulative intents well, others did not

shots from the questionnaire. It became apparent that some children have a basic understanding of the manipulative power and purpose behind specific design elements when encountering dark patterns, while every third child has not. Furthermore, it was evident that for some types of dark patterns, it was easier to make this connection than for other types. For example, children were more confident in recognising aesthetic manipulations compared to linguistic tricks or covert tactics. However, we were unable to determine a clear connection between our observations and children's prior exposure to dark patterns. Therefore, the question of the influence of dark pattern prevalence on children's mental models remains open for further exploration.

Some dark pattern types were more easily comprehensible than others

Our findings on children's mental models of dark patterns help identify approaches for future research on countermeasures to protect children from dark patterns. Countermeasures discussed in the past for adults, such as laws and regulations, establishing ethical designs, and education, could also be effectively applied in an adapted form in the context of children. Further research is needed to effectively determine the appropriate adaptations. Possible approaches for future work are discussed below.

Future work is needed to develop effective countermeasures

5.2 Future Work

One direction for future work would be to build on our thesis and extend the knowledge about children's mental models of dark patterns. The limitations of our study described in Section 4.6 "Limitations" could be rectified in a related study. For example, our study could be repeated using a comparable questionnaire that focuses on other types of dark patterns. Insights into the differences in mental models for varying patterns would allow the design of countermeasures that are more targeted and customised towards the manipulation. Furthermore, the influence of the prevalence of dark patterns in children's everyday lives could be investigated more deeply in a dedicated, more focused study.

Our study could be repeated in an adapted manner to resolve discussed limitations

In order to identify the exact sources of danger and to be able to design countermeasures accordingly, it would also be interesting to study children in their interaction with dark patterns. In particular, the topic of preference inconsistency should be explored more closely. On the one hand, one could look at how children react to dark patterns if they are only exposed to them unconsciously instead of actively looking for manipulative elements. Will they still recognise dark patterns? Are they influenced in their decisions by the presence of dark patterns? Such insights might introduce promising directions for further countermeasure development, as unconscious encounters with dark patterns match a realistic scenario more closely. On the other hand, it is necessary to investigate how perception and behaviour differ when encountering different dark pattern types. Which dark patterns are recognised and for which is the preference inconsistency of children particularly high? Which are perceived as dangerous and which are given in to, possibly out of convenience or annoyance?

It is necessary to investigate preference inconsistency depending on unconscious exposure and dark pattern type

Once the research field has acquired a better understanding of children in the context of dark patterns, different countermeasure approaches should be designed, tested and refined. One approach would be to evaluate the effectiveness of educational interventions. How do children's mental models change after being educated about dark patterns? How does this reflect in how they perceive and interact with dark patterns? A longitudinal study could also investigate the impact of different educational techniques and the level of education on children's behaviour and, in particular, preference inconsistency. For example, one could distinguish between theoretical teaching sessions in school, practical exercises with parents, and autonomous training through a gamified learning app, and measure the effectiveness of each method. It would also be interesting to assess whether children are more likely to resist manipulation if they have been informed of the possible risks of dark patterns. This information could be used to develop appropriate educational interventions which could be used to sustainably protect children from dark patterns. Altogether, there are still many open questions, but also new and promising research approaches in the field of children and dark patterns, which are essential to be tackled, especially in a world where online manipulation is prevalent.

Evaluating the effectiveness of educational interventions could help to develop countermeasures

Appendix A

A Collection of Dark Pattern Types

In the following, we provide a collection of dark pattern types we used in the questionnaire on the analysis of the results, taken from the taxonomies mentioned in 2.1.1 "Taxonomies". It contains patterns from the works of Brignull et al. [2010], Gray et al. [2018], and Mathur et al. [2019].

Variant	Description	Source
Visual Interference	Using style and visual presentation to steer users to or away from certain choices	*Mathur et al. [2019]*
Toying with Emotions	Emotionally manipulative framing	*Gray et al. [2018]*
Confirmshaming	Choice framed in a way that makes it seem dishonorable, stupid	*Brignull et al. [2010], Mathur et al. [2019]*
Trick Questions	Intentional or obvious ambiguity	*Gray et al. [2018], Mathur et al. [2019]*
Preselection	Firm-friendly default is preselected	*Gray et al. [2018]*

Figure A.1: A collection of dark pattern types mentioned in this thesis. The patterns originate from the taxonomies of Brignull et al. [2010], Gray et al. [2018], Mathur et al. [2019]. Table adapted from Luguri and Strahilevitz [2021] and Mathur et al. [2019].

Appendix B

Mental Models Questionnaire

The questionnaire we used for our mental models study with three classes of fifth-graders is attached in the following. An English translation of the German original is provided right after in B.2 "Mental Models Questionnaire (English)". Note that, due to counterbalancing methods applied, four versions of the same questionnaire with different orderings of the stimulus material in Part 1 were used in the study. We deliberately omit to attach all four versions here.

B.1 Mental Models Questionnaire (German)

The original, German version of the questionnaire is given hereafter:

Masterarbeitsstudie Medienkompetenz von Kindern

Sarah Sahabi

Zustimmung

Wir führen eine Studie zum Thema "Medienkompetenz von Kindern" durch. Die Studie dauert etwa 90 Minuten mit einer kurzen Pause. Wenn du einverstanden bist und an der Studie teilnehmen möchtest, wirst du anonym Fragebögen ausfüllen und durch Diskussionen in der Gruppe mehr zu dem Thema lernen.
Um teilnehmen zu können, musst du eine unterschriebene Einverständniserklärung eines Elternteils abgegeben haben. Deine Teilnahme ist freiwillig und du darfst jederzeit aufhören oder nach einer Pause fragen. Wenn du mit der Teilnahme einverstanden bist, kreuze bitte die folgende Box an:

☐ Ich habe die Informationen gelesen und möchte an der Studie teilnehmen.

Danke für deine Unterstützung!

Persönliche Angaben

Bitte fülle die folgenden Fragen über dich aus. Die Informationen werden anonym gesammelt.

Alter:

Geschlecht:

Wie häufig benutzt du ungefähr ein Smartphone oder Tablet? Kreuze an, was am besten passt:

- ☐ Mindestens einmal am Tag
- ☐ Mehrfach pro Woche
- ☐ Höchstens einmal pro Woche
- ☐ Höchstens 1-2 mal pro Monat
- ☐ Gar nicht

Bitte gib an, welche Apps du sehr häufig verwendest:

Teil 1

Bei dem Spiel "4 Jagd" erhältst du bei jedem Zug jeweils vier Spielkarten mit leicht verschiedenen Bildern einer App. Bei deinem ersten Zug erhältst du die folgenden vier Bilder einer Spiele App, bei der alle Leben aufgebraucht wurden. Um wieder neue Leben zu bekommen und weiterspielen zu können, muss man entweder warten, oder sich eine Werbung ansehen, nach der man direkt weiterspielen kann.

Sieh dir die vier verschiedenen Bilder an und überlege: Wie schön, kompliziert, und bedrohlich findest du die Apps auf den Bildern? Ordne die vier Bilder jeweils in die untenstehenden Strahlen ein, wie in dem Beispiel gezeigt.

Nicht schön — Neutral — Sehr schön

Nicht kompliziert — Neutral — Sehr kompliziert

Nicht vertrauenswürdig — Neutral — Sehr vertrauenswürdig

Teil 2.1

Sieh dir die beiden folgenden Bilder genau an. Erkennst du die Unterschiede zwischen den beiden Bildern? Markiere und beschreibe die Unterschiede in der Box unter den Bildern.

Beschreibe die Unterschiede hier:

Teil 2.2

Nachdem du deine Spielkarten bekommen und genau angesehen hast, bekommst du außerdem bei jedem Zug ein geheimes Ziel, das du erreichen musst, um das Spiel *4 Jagd* zu gewinnen. Um diese Runde zu gewinnen, musst du folgendes Ziel erreichen: Wähle das Bild aus, mit dem du die meisten Personen dazu bringst, sich einen Werbespot anzusehen. Dafür kannst du zwischen den beiden Bildern oben auf dieser Seite wählen. Welches Bild wählst du? Kreuze an und begründe deine Entscheidung in der folgenden Box.

Ich wähle dieses Bild:

- ☐ Bild U5
- ☐ Bild G6

Begründung:

Teil 2.3

Nun ist eine andere Mitspielerin am Zug. Sie zieht ihre Bilder und hat schließlich die Wahl zwischen den folgenden beiden Bildern einer Webseite, die Cookies verwendet. Dabei kann man entweder akzeptieren oder ablehnen, dass alle Cookies verwendet werden dürfen. Nach längerem Überlegen entscheidet sich deine Mitspielerin für das untere Bild, Bild T3. Ausgehend von ihrer Entscheidung, was könnte ihr geheimes Ziel gewesen sein, das sie erreichen möchte? Beschreibe das Ziel in der Box unten auf der Seite und erkläre, wie das Bild T3 dieses Ziel erfüllt.
Tipp: Sieh dir, wie in Teil 2.1, zuerst genau die Unterschiede zwischen den beiden Bildern an.

Bitte beschreibe und begründe deine Antwort für Teil 2.3 in der folgenden Box:

Freiwillige Bonusaufgabe: wie könnte man dieses Ziel noch erreichen? Überlege dir eine zweite Möglichkeit, wie man die Seite von dem Bild F7 verändern könnte, um das beschriebene Ziel zu erreichen. Male deine Idee als Bild in die folgende Box, und vergiss nicht, danach dein Bild in der Box darunter zu beschreiben.

Bitte beschreibe in der folgenden Box:

Teil 3

Auf dem folgenden Bild möchte ein Spieler die Spiele App schließen. Bevor sich die App schließt, wird der Spieler noch einmal gefragt, ob er sich sicher ist, dass er nicht weiterspielen möchte und ob er seine erreichte Punktzahl gerne mit seinen Freunden teilen möchte. Sieh dir das Bild an, denn jetzt bist du dran: In das Bild haben sich vier solcher Manipulationen eingeschlichen. Kannst du sie entdecken? Kreise alle Manipulationen, die du finden kannst, direkt im Bild ein und begründe deine Entscheidungen in der Box darunter.

Bitte erkläre deine Entscheidungen in der folgenden Box:

B.2 Mental Models Questionnaire (English)

An English translation of the German version of the questionnaire is given in the following:

Master's thesis study on children's media competence

Sarah Sahabi

Consent

We are conducting a study on the topic of children's media literacy. The study will last about 90 minutes with a short break. If you agree and want to participate in the study, you will anonymously fill in questionnaires and learn more about the topic through group discussions.
To participate, you must have a signed consent form from a parent. Your participation is voluntary and you may stop or ask for a break at any time. If you agree to participate, please tick the box below:

☐ I have read the information and would like to participate in the study.

Thank you for your support!

Personal details

Please complete the following questions about yourself. The information will be collected anonymously.

Age:

Gender:

How often do you use a smartphone or tablet? Tick what fits best:

☐ At least once a day

☐ Several times a week

☐ At most once a week

☐ At most 1-2 times per month

☐ Not at all

Please indicate which apps you use frequently:

Part 1

In the game "4 Hunt", you receive four game cards with slightly different pictures of an app on each turn. On your first turn, you receive the following four pictures of a game app where all lives have been used up. To get new lives again and continue playing, you either have to wait or watch a commercial, after which you can continue playing directly.

Look at the four different pictures and think: How beautiful, complicated and threatening do you find the apps in the pictures? Put each of the four pictures into the rays below, as shown in the example.

Not beautiful — Neutral — Very beautiful

Not complex — Neutral — Very complex

Not trustworthy — Neutral — Very trustworthy

Part 2.1

Look carefully at the two pictures below. Can you see the differences between the two pictures? Mark and describe the differences in the box below the pictures.

Describe the differences here:

Part 2.2

After you have received your playing cards and looked at them carefully, you will also be given a secret goal each turn that you must achieve in order to win the game 4 Hunt. To win this round, you must achieve the following goal: Choose the picture that will get the most people to watch a commercial. You can choose between the two pictures at the top of this page. Which picture do you choose? Tick the box and give reasons for your decision in the box below.

I choose this image:

- ☐ Image U5
- ☐ Image G6

Justification:

Part 2.3

Now it is another player's turn. She draws her pictures and finally has the choice between the following two pictures of a website that uses cookies. You can either accept or reject that all cookies may be used. After thinking about it for a while, your teammate decides on the lower picture, picture T3.
Based on her decision, what could have been her secret goal that she wants to achieve? Describe the goal in the box at the bottom of the page and explain how picture T3 fulfils this goal.
Tip: As in part 2.1, first look carefully at the differences between the two pictures.

Please describe and justify your answer for part 2.3 in the box below:

Voluntary bonus task: how else could you achieve this goal? Think of a second way to change the side of picture F7 to achieve the goal described. Draw your idea as a picture in the box below, and don't forget to describe your picture in the box below.

Please describe in the box below:

Part 3

In the following picture, a player wants to close the games app. Before the app closes, the player is asked once again if he is sure he does not want to continue playing and if he would like to share his score with his friends. Look at the picture, because now it's your turn: four such manipulations have crept into the picture. Can you spot them? Circle all the manipulations you can find directly in the picture and give reasons for your decisions in the box below.

Please explain your choices in the box below:

Appendix C

Most Frequently Used Apps and Dark Patterns They Contain

The following table lists all apps the participants stated to be using frequently, together with the dark patterns they contain according to Di Geronimo et al. [2020] and Lupiáñez-Villanueva et al. [2022].

App Category	App	Frequency	Contained Dark Patterns
Communication	Whatsapp	62.1 %	Nagging, Roach Motel, Visual Interference, Preselection
	Teams	43.9 %	
	Telefon	16.7 %	
	Discord	3.0 %	Roach Motel, Preselection, Privacy Zuckerig
	Signal	3.0 %	Nagging, Roach Motel, Preselection, Visual Interference
	Telegram	1.5 %	Roach Motel, Preselection
	Threema	1.5 %	
	Viber	1.5 %	Roach Motel, Hidden Information, Preselection, Disguised Ad, Forced Action, Privacy Zuckerig
Entertainment	YouTube	56.1 %	Roach Motel, Nagging, Preselection, Forced Registration, Disguised Ad, Hidden Information
	TikTok	16.7 %	Disguised Ad, Intermediate Currency, Toying With Emotion, Hidden Information, Preselection, Privacy Zuckerig
	Fotos	15.2 %	
	Spotify	12.1 %	Forced Continuity, Nagging, Roach Motel, Preselection
	Snapchat	10.6 %	Nagging, Roach Motel, Preselection, Toying With Emotion, Visual Interference, Intermediate Currency
	Netflix	9.1 %	Roach Motel, Preselection, Forced Registration, Toying With Emotions, Confirmshaming
	Instagram	4.5 %	Nagging, Roach Motel, Hidden Information, Preselection, Toying With Emotion, Visual Interference, Privacy Zuckerig, Disguised Ad, Forced Registration
	Deezer	1.5 %	Nagging, Roach Motel, Preselection, Visual Interference
Browser	Google	25.8 %	Disguised Ad, Visual Interference, Forced Registration, Toying With Emotions, Preselection, Roach Motel
	Safari	9.1 %	Disguised Ad, Visual Interference, Forced Registration, Toying With Emotions, Preselection, Roach Motel
Games	Roblox	18.2 %	Intermediate Currency, Roach Motel, Preselection, Visual Interference
	Weitere Spiele	16.7 %	Intermediate Currency, Nagging, Forced Registration, Countdown Timer, Visual Interference, Roach Motel, Preselection, Bait and Switch
	Brawl Stars	10.6 %	Intermediate Currency

	Clash of Clans	3.0 %	
	Fortnite	3.0 %	
	Hogwarts Mystery	3.0 %	
	More games	10.6 %	
Education	Anton	15.2 %	(no access)
	Phase 6	3.0 %	
	Duolingo	3.0 %	
	Schlaukopf	3.0 %	
Utility	Emojiapp	3.0 %	
	J YouPro	3.0 %	
	Malapp	3.0 %	
	Muslim Pro	3.0 %	
	Notizen	3.0 %	
	Rechner	3.0 %	
	Übersetzer	3.0 %	
	Video Editor	3.0 %	
	Busapp	3.0 %	
News	Not specified	3.0 %	Forced Registration, Forced Continuity, Roach Motel, Visual Interference, Preselection, Toying With Emotions, Nagging, Hidden Information, Disguised Ad, Privacy Zuckerig

Figure C.1: All frequently used apps the participants mentioned and their frequencies, categorised in app categories. The table also contains a list of dark patterns contained in each app, as collected by Di Geronimo et al. [2020] and Lupiáñez-Villanueva et al. [2022].

Appendix D

Codebook

The following codebook contains all codes and categories derived during the qualitative analyses of Part 2.1 to Part 3 of the mental models questionnaire.

Code System	Frequency
All Codes	**697**
Part 2.1	**15**
Alternative not seen	10
Part 2.2	**105**
Visual Interference detected	28
Bad close option visibility	20
Ad-option eye-catching	8
Close option harder to reach	6
G6 has less text	2
G6 has no close option	20
U5 is more trustworthy	3
Task not understood	12
Part 2.3	**222**
Goals	45
Accept cookies	44
Data privacy	1
Justification	34
Button size	20
Button visibility	18
Button colour	5
Faster to accept	4
Accept seems better	1
No reject option	1
Trustworthiness	1
Differences found without goal	16
Differences found	9
Decline button visibility	6
No decline option	5
Security	1
Bonus Part	**165**
Visual Interference	29
Accept bigger	14
Accept more colourful	14
Decline greyed out	7
Decline position unexpected	4
Alternative hidden	13
Undesirable alternative	9
Compromise alternative	2
No alternative	6

	Compromise alternative	2
	No alternative	6
	Alternative: close	1
	Alternative: wait	1
Toying with Emotion		13
	Influential formulation	7
	Emojis/Symbols	6
	Colours	3
Bait and Switch		3
Bribery		2
Confusing Terminology		1
Trustworthiness		1
Unclear		7
Task not understood		2
Part 3		**189**
Toying with Emotions		39
	Colour	24
	Shaming to play on	8
	Cookie banner formulation to play on	1
	Cookie banner formulation to accept	5
	Unsettling questioning	15
Visual Interference		7
	Button size	7
	Button order	1
Confirmshaming		14
Preselection		3
Trick Question		25
Mistrust		16
	Misunderstood cookie banner	8
	Privacy concerns share points	4

Figure D.1: Codebook containing all codes and categories from the qualitative analyses of parts 2.1–3, together with the number of segments that were assigned each code.

Bibliography

Alice Bell. Designing and testing questionnaires for children. *Journal of Research in Nursing*, 12(5):461–469, 2007. doi:10.1177/1744987107079616.

Beth T. Bell and Daniel Fitton. Dark Patterns in Mobile Games: A Source of Online Risk for Youths? *What Can CHI Do About Dark Patterns? CHI WORKSHOP*, 2021. URL `https://drive.google.com/file/d/1Y5hbXyc1QAwaSSICcj8NJWG8eLgBv2zo/view`.

Orit Ben-Zvi Assaraf, Haim Eshach, Nir Orion, and Yousif Alamour. Cultural differences and students' spontaneous models of the water cycle: A case study of jewish and bedouin children in israel. *Cultural Studies of Science Education*, 7:451–477, 2012. doi:10.1007/s11422-012-9391-5.

Sorin Berbece. Let there be light! dark patterns under the lens of the eu legal framework. Available at SSRN, 2019.

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*, pages 763–776, 2021. doi:10.1145/3461778.3462086.

Ann Bostrom. Lead is like mercury: risk comparisons, analogies and mental models. *Journal of Risk Research*, 11(1-2):99–117, 2008. ISSN 1366-9877. doi:10.1080/13669870701602956.

Jo Boyden and Judith Ennew. *Children in focus: a manual for participatory research with children*. Save the Children Sweden, 1997.

Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3 (2):77–101, 2006. doi:10.1191/1478088706qp063oa. URL `https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa`.

Virginia Braun, Victoria Clarke, Nikki Hayfield, and Gareth Terry. *Thematic Analysis*, pages 843–860. Springer Singapore, Singapore, 2019. ISBN 978-981-10-5251-4. doi:10.1007/978-981-10-5251-4_103. URL `https://doi.org/10.1007/978-981-10-5251-4_103`.

Harry Brignull, Mark Leiser, Cristiana Santos, and Kosha Doshi. Deceptive patterns – user interfaces designed to trick you, 2010. URL `https://www.deceptive.design/`. *Accessed: May 12, 2023*. Note: The website underwent a major update on April 25, 2023 (now archived under `https://www.old.deceptive.design/`). We refer to the current version.

Jessica E. Brodsky, Arshia K. Lodhi, Kasey L. Powers, Fran C. Blumberg, and Patricia J. Brooks. "it's just everywhere now": Middle-school and college students' mental models of the internet. *Human Behavior and Emerging Technologies*, 3(4):495–511, 2021. ISSN 2578-1863. doi:10.1002/hbe2.281.

Amy Bruckman, Alisa Bandlow, Jill Dimond, and Andrea Forte. *Human–computer interaction for kids*, pages 841–861. CRC Press, 2012.

Sawssen Garbouj Chaouachi and Kaouther Saied Ben Rached. Perceived deception in advertising: Proposition of a measurement scale. *Journal of Marketing research & Case studies*, page 1, 2012. doi:10.5171/2012.712622.

Seung Youn (Yonnie) Chyung, Ieva Swanson, Katherine Roberts, and Andrea Hankinson. Evidence-based survey design: The use of continuous rating scales in surveys. *Performance Improvement*, 57(5):38–48, 2018. doi:10.1002/pfi.21763.

Bryan Clift, Kovila Coopamootoo, Cigdem Sengul, Karen Renaud, and Ben Morrison. Methodology: Measuring Young Learners' Mental Models of Online Dark Patterns (Sludge), 2022.

Gregory Conti and Edward Sobiesk. *Malicious Interface Design: Exploiting the User*. WWW '10. Association for Computing Machinery, New York, NY, USA, 2010. ISBN 9781605587998. doi:10.1145/1772690.1772719.

June Cotte, Robin A. Coulter, and Melissa Moore. Enhancing or disrupting guilt: the role of ad credibility and perceived manipulative intent. *Journal of Business Research*, 58(3):361–368, 2005. ISSN 01482963. doi:10.1016/S0148-2963(03)00102-4. Special Section: Marketing Communications and Consumer Behavior.

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, 2021.

Pearl Denham. Nine- to fourteen-year-old children's conception of computers using drawings. *Behaviour & Information Technology*, 12(6):346–358, 1993. doi:10.1080/01449299308924399.

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, 2020. Association for Computing Machinery. ISBN 9781450367080. doi:10.1145/3313831.3376600. URL `https://doi.org/10.1145/3313831.3376600`.

Jérôme Dinet and Muneo Kitajima. *Draw me the Web. Impact of mental model of the Web on information search performance of young users*. IHM '11. Association for Computing Machinery, New York, NY, USA, 2011. ISBN 9781450308229. doi:10.1145/2044354.2044358.

Emma E.H. Doyle, Sara E. Harrison, Stephen R. Hill, Matt Williams, Douglas Paton, and Ann Bostrom. Eliciting mental models of science and risk for disaster communication: A scoping review of methodologies. *International Journal of Disaster Risk Reduction*, 77:103084, 2022. ISSN 2212-4209. doi:10.1016/j.ijdrr.2022.103084.

James K. Doyle and David N. Ford. Mental models concepts for system dynamics research. *System Dynamics Review*, 14(1):3–29, 1998. ISSN 0883-7066. doi:10.1002/(SICI)1099-1727(199821)14:1<3::AID-SDR140>3.0.CO;2-K.

Allison Druin. Cooperative inquiry: Developing new technologies for children with children. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, pages 592—599, New York, NY, USA, 1999. Association for Computing Machinery. ISBN 0201485591. doi:10.1145/302979.303166.

Linn Ekroth and Josefine Sandqvist. Confirmshaming and its effect on users: A qualitative study on how confirmshaming in unsubscription processes affect users, 2020.

Gary A. Fine and Kent L. Sandstrom. *Knowing children: Participant observation with minors*, volume 15. SAGE Publications, Incorporated, 1988. doi:10.4135/9781412984706.

Dan Fitton and Janet C. Read. Creating a framework to support the critical consideration of dark design aspects in free-to-play apps. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*, IDC '19, page 407–418, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450366908. doi:10.1145/3311927.3323136.

Dedre Gentner and Donald R. Gentner. Flowing waters or teeming crowds: Mental models of electricity. In *Mental models*, pages 107–138. Psychology Press, 2014.

Shawn Glynn. Drawing mental models. *The Science Teacher*, 64(1):30, 1997.

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, page 1–14, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450356206. doi:10.1145/3173574.3174108.

Robin S. Grenier and Dana Dudzinska-Przesmitzki. A conceptual model for eliciting mental models using a composite methodology. *Human Resource Development Review*, 14(2):163–184, 2015. ISSN 1534-4843. doi:10.1177/1534484315575966.

Jens Grossklags and Alessandro Acquisti. When 25 cents is too much: An experiment on willingness-to-sell and willingness-to-protect personal information. In *WEIS*, 2007.

Johanna Gunawan, Amogh Pradeep, David Choffnes, Woodrow Hartzog, and Christo Wilson. A comparative study of dark patterns across web and mobile modalities. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2), 2021. doi:10.1145/3479521.

Philip Hausner and Michael Gertz. Dark patterns in the interaction with cookie banners. In Regan Mandryk, Mark Hancock, Mark Perry, and Anna Cox, editors, *What Can CHI Do About Dark Patterns? CHI WORKSHOP*, pages 1–14, New York, NY, USA, 2021. ACM. ISBN 9781450356206. doi:10.48550/arXiv.2103.14956.

Luiza Jarovsky. Dark patterns in personal data collection: Definition, taxonomy and lawfulness. Available at SSRN, 2022.

P.N. Johnson-Laird. Mental models in cognitive science. *Cognitive Science*, 4(1):71–115, 1980. ISSN 0364-0213. doi:https://doi.org/10.1016/S0364-0213(81)80005-5.

Natalie A. Jones, Helen Ross, Timothy Lynam, Pascal Perez, and Anne Leitch. Mental models: An interdisciplinary synthesis of theory and methods. *Ecology and society*, 16(1), 2011. URL `http://www.jstor.org/stable/26268859`.

Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. My data just goes everywhere: User mental models of the internet and implications for privacy and security. In *Proceedings of the Eleventh USENIX Conference on Usable Privacy and Security*, SOUPS '15, page 39–52, USA, 2015. USENIX Association. ISBN 9781931971249.

Christie Kodama, Beth St. Jean, Mega Subramaniam, and Natalie Greene Taylor. There's a creepy guy on the other end at Google!: engaging middle school students in a drawing activity to elicit their mental models of Google. *Information Retrieval Journal*, 20(5):403–432, 2017. doi:10.1007/s10791-017-9306-x.

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, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380959. doi:10.1145/3411763.3451632.

Priya Kumar, Shalmali Milind Naik, Utkarsha Ramesh Devkar, Marshini Chetty, Tamara L. Clegg, and Jessica Vitak. No telling passcodes out because they're private: Understanding children's mental models of privacy and security online. *Proc. ACM Hum.-Comput. Interact.*, 1 (CSCW), dec 2017. doi:10.1145/3134699.

Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research Methods in Human Computer Interaction (Second Edition)*. Morgan Kaufmann, 2017. ISBN 978-0-12-805390-4. URL `https://www.sciencedirect.com/book/9780128053904/research-methods-in-human-computer-interaction`.

Barry M. Leiner, Vinton G. Cerf, David D. Clark, Robert E. Kahn, Leonard Kleinrock, Daniel C. Lynch, Jon Postel, Larry G. Roberts, and Stephen Wolff. A brief history of the internet. *ACM SIGCOMM Computer Communication Review*, 39(5):22–31, October 2009. ISSN 0146-4833. doi:10.1145/1629607.1629613.

Gitte Lindgaard, Cathy Dudek, Devjani Sen, Livia Sumegi, and Patrick Noonan. An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages. *ACM Trans. Comput.-Hum. Interact.*, 18(1), may 2011. ISSN 1073-0516. doi:10.1145/1959022.1959023.

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

Francisco Lupiáñez-Villanueva, Alba Boluda, Francesco Bogliacino, Giovanni Liva, Lucie Lechardoy, and Teresa Rodríguez de las Heras Ballell. *Behavioural study on unfair commercial practices in the digital environment: dark patterns and manipulative personalisation*. Publications Office of the European Union, 2022. doi:10.2838/859030.

Ana Maria Marhan, Mihai Ioan Micle, Camelia Popa, and Georgeta Preda. A review of mental models research in child-computer interaction. *Procedia-Social and Behavioral Sciences*, 33:368–372, 2012. doi:10.1016/j.sbspro.2012.01.145.

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. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), nov 2019. doi: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, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380966. doi:10.1145/3411764.3445610.

Philipp Mayring. Qualitative content analysis: theoretical foundation, basic procedures and software solution, 2014.

Karen Mcinnes, Justine Howard, Gareth Miles, and Kevin Crowley. Behavioural differences exhibited by children when practising a task under formal and playful conditions. *Educational and Child Psychology*, 26:31–39, 01 2009. doi:10.53841/bpsecp.2009.26.2.31.

Pekka Mertala. Using playful methods to understand children's digital literacies. In *Research Through Play: Participatory Methods in Early Childhood*. Sage, 05 2021.

Thomas Mildner and Gian-Luca Savino. Ethical user interfaces: Exploring the effects of dark patterns on facebook, 2021.

Aliaksei Miniukovich and Antonella De Angeli. Quantification of interface visual complexity. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*, AVI '14, page 153–160, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450327756. doi:10.1145/2598153.2598173.

Tobias Mocka. *Der Einfluss von Dark Patterns auf die Usability und User Experience*. PhD thesis, Technische Hochschule Ingolstadt, 2022.

Ben Morrison, Cigdem Sengul, Mark Springett, Jacqui Taylor, and Karen Renaud. White Paper: Mental Models of Dark Patterns, 2021. URL `http://eprints.bournemouth.ac.uk/37008/`.

Arvind Narayanan, Arunesh Mathur, Marshini Chetty, and Mihir Kshirsagar. Dark patterns: Past, present, and future: The evolution of tricky user interfaces. *Queue*, 18(2):67–92, May 2020. ISSN 1542-7730. doi:10.1145/3400899.3400901.

Donald A. Norman. Some Observations on Mental Models. *Mental Models*, 1983.

Donald A. Norman. *The Design of Everyday Things*. Basic Books, Inc., USA, 2002. ISBN 9780465067107.

Maggie Oates, Yama Ahmadullah, Abigail Marsh, Chelse Swoopes, Shikun Zhang, Rebecca Balebako, and Lorrie Faith Cranor. Turtles, Locks, and Bathrooms: Understanding Mental Models of Privacy Through Illustration. *Proceedings on Privacy Enhancing Technologies*, 2018(4):5–32, 2018. doi:10.1515/popets-2018-0029.

Georgia Panagiotaki, Gavin Nobes, and Robin Banerjee. Children's representations of the earth: A methodological comparison. *British Journal of Developmental Psychology*, 24(2):353–372, 2006. doi:10.1348/026151005X39116.

Nils Pancratz and Ira Diethelm. Draw us how smartphones, video gaming consoles, and robotic vacuum cleaners look like from the inside: Students' conceptions of computing system architecture. In *Proceedings of the*

15th Workshop on Primary and Secondary Computing Education, WiPSCE '20, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450387590. doi:10.1145/3421590.3421600.

Marina Papastergiou. Students' mental models of the internet and their didactical exploitation in informatics education. *Education and Information Technologies*, 10:341–360, 2005.

P. J. Pridmore and R. G. Lansdown. Exploring Children's Perceptions of Health: Does Drawing Really Break Down Barriers? *Health Education Journal*, 56(3):219–230, 1997. doi:10.1177/001789699705600302.

Pavol Prokop, Jana Fančovičová, and Sue Dale Tunnicliffe. The Effect of Type of Instruction on Expression of Children's Knowledge: How Do Children See the Endocrine and Urinary System? *International Journal of Environmental and Science Education*, 4(1):75–93, 2008.

Samantha Punch. Research with children: The same or different from research with adults? *Childhood*, 9(3):321–341, 2002. ISSN 0907-5682. doi:10.1177/0907568202009003005.

Bernhard Rohleder. Bitcom Kinder- und Jugendstudie 2022, 2022. URL `https://www.bitkom.org/sites/main/files/2022-06/Bitkom-Charts_Kinder_Jugendliche_09.06.2022_0.pdf`.

Johnny Saldaña. *The coding manual for qualitative researchers*. SAGE publications Ltd, 2013.

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. *arXiv e-prints*, pages arXiv–2204, 2022. doi:10.48550/arXiv.2204.11836.

Nancy Staggers and A.F. Norcio. Mental models: concepts for human-computer interaction research. *International Journal of Man-Machine Studies*, 38(4):587–605, 1993. ISSN 0020-7373. doi:https://doi.org/10.1006/imms.1993.1028. URL `https://www.sciencedirect.com/science/article/pii/S002073738371028X`.

Andrew Thatcher and Mike Greyling. Mental models of the internet. *International Journal of Industrial Ergonomics*, 22(4):299–305, 1998. ISSN 0169-8141. doi:10.1016/S0169-8141(97)00081-4.

Patti M. Valkenburg and Jessica T. Piotrowski. *Plugged In: How Media Attract and Affect Youth*. Yale University Press, 2017. ISBN 9780300218879. doi:10.12987/yale/9780300218879.001.0001.

Christof van Nimwegen and Jesse de Wit. Shopping in the dark: Effects of platform choice on dark pattern recognition. In Masaaki Kurosu, editor, *Human-Computer Interaction. User Experience and Behavior*, volume 13304 of *Lecture Notes in Computer Science*, pages 462–475. Springer International Publishing, Berlin, Heidelberg, 2022. ISBN 978-3-031-05411-2. doi:10.1007/978-3-031-05412-9_32.

Christian Voigt, Stephan Schlögl, and Aleksander Groth. Dark patterns in online shopping: of sneaky tricks, perceived annoyance and respective brand trust. In Fiona Fui-Hoon Nah and Keng Siau, editors, *HCI in Business, Government and Organizations*, pages 143–155, Cham, 2021. Springer International Publishing. ISBN 978-3-030-77750-0.

Quirin Weinzierl. Dark patterns als herausforderung für das recht: Rechtlicher schutz vor der ausnutzung von verhaltensanomalien, 2020. ISSN 0721-880X.

Barry Wellman, Anabel Quan-Haase, Jeffrey Boase, Wenhong Chen, Keith Hampton, Isabel Díaz, and Kakuko Miyata. The Social Affordances of the Internet for Networked Individualism. *Journal of Computer-Mediated Communication*, 8(3):JCMC834, 04 2003. ISSN 1083-6101. doi:10.1111/j.1083-6101.2003.tb00216.x.

Yaxing Yao, Davide Lo Re, and Yang Wang. Folk Models of Online Behavioral Advertising. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW ’17, pages 1957–1969, New York, NY, USA, 2017. Association for Computing Machinery. doi:10.1145/2998181.2998316.

Jason C. Yip, Kiley Sobel, Xin Gao, Allison Marie Hishikawa, Alexis Lim, Laura Meng, Romaine Flor Ofiana, Justin Park, and Alexis Hiniker. Laughing is scary, but farting is cute: A conceptual model of children's perspectives of creepy technologies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–15, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450359702. doi:10.1145/3290605.3300303.

José P. Zagal, Staffan Björk, and Chris Lewis. Dark patterns in the design of games. In *Foundations of Digital Games 2013*, 2013.

Yan Zhang. The influence of mental models on undergraduate students' searching behavior on the web. *Information Processing & Management*, 44(3):1330–1345, 2008. ISSN 0306-4573. doi:https://doi.org/10.1016/j.ipm.2007.09.002. URL `https://www.sciencedirect.com/science/article/pii/S0306457307001732`.

Leah Zhang-Kennedy, Christine Mekhail, Yomna Abdelaziz, and Sonia Chiasson. From nosy little brothers to stranger-danger: Children and parents' perception of mobile threats. In *Proceedings of the The 15th International Conference on Interaction Design and Children*, IDC '16, page 388–399, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450343138. doi:10.1145/2930674.2930716. URL `https://doi.org/10.1145/2930674.2930716`.

Jun Zhao, Ge Wang, Carys Dally, Petr Slovak, Julian Edbrooke-Childs, Max Van Kleek, and Nigel Shadbolt. 'i make up a silly name': Understanding children's perception of privacy risks online. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–13, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450359702. doi:10.1145/3290605.3300336.

Index

Typeset July 7, 2023