WHITE PAPER:
Mental Models of Dark Patterns

Benjamin Morrison[1], Cigdem Sengul[2], Mark Springett[3], Jacqui Taylor[4], Karen Renaud[5]
[1]Northumbria University (benjamin.a.morrison@northumbria.ac.uk)
[2]Brunel University (cigdem.sengul@brunel.ac.uk)
[3]Middlesex University (m.springett@mdx.ac.uk)
[4]Bournemouth University (jtaylor@bournemouth.ac.uk)
[5]University of Strathclyde (karen.renaud@strath.ac.uk)

November 2021

Abstract

In this paper, we report on the literature related to understanding young learners' mental models related to deceptive "dark patterns" used by malicious agents online: so-called *sludge*. We also discuss elicitation of mental models, particularly when carrying out activities to reveal the mental models of young learners. In addition, we review the ethical considerations when carrying out research in this domain. Finally, we propose the design of an activity to implement the lessons we have learned to assess the sludge-related mental models of young learners.

1 Introduction

The use of persuasive techniques, or nudging, is an established strategy used in several areas of online and offline activity. It is used in e-commerce, e-health and public communication, for example. Sometimes health apps are gamified, or user interfaces tweaked to strategically present information in a particular way to influence decisions. However, while the authors of the initial book that introduced nudging to the world [76] always sign their books with "*Nudge for Good*", nudging is sometimes undeniably also used for nefarious purposes [52, 61]. In these cases, the use of influential techniques is called 'sludge'.

The cynical use of influential techniques presents an increasing threat to the well-being of users, especially in the case of underage users. In 2018, the United Nations Conference on Trade and Development referred to *sludge* [81] i.e., harmful nudging, as a concerning development in online commerce: "*Consumers may not be made aware of particular factors influencing their decisions, or realise they are being manipulated in ways they don't understand, leading them to take decisions they otherwise would not have.*".

Children are among the groups cited by the report as being particularly 'at risk', along with those without 'technological savvy'. The EU Kids Online report [44] highlights that children are spending more time online, accessing more online material, in more diverse ways and at younger ages. 53% of children own a smartphone by the age of 11 years, with that number rising to 84% among teenagers [67]. Social media is a key area where sludge is popularly used, and young learners might well be vulnerable to persuasion which is not 'for good'. Teenagers are seen as vulnerable due to their limited capacity for

self-regulation and a particular susceptibility to peer pressure [3]. Moreover, teens may be inclined to treat risky behaviour as a learning experience [87].

Hamdan *et al.* [32] identify the types of online harm that are pertinent to teenagers, which include: (1) privacy breaches for commercial exploitation, (2) cyberbullying and (3) cyberstalking. Some approaches to security emphasise parental control, but the effectiveness of this is often undermined by parents' lack of expertise [53] and by teens' reliance on the Internet for information [36], which might lead to their being misinformed. A further issue is that teenagers often hide information from their parents whilst sharing much of their lives with their social connections [87]. Whilst the assumption that teenagers' awareness of privacy issues is weaker than that of adults has been challenged in recent research [12], there is indeed evidence that their actual online behaviours often do not demonstrate heightened awareness and lead to greater vulnerability.

Developmental psychologists have traditionally used seven stages to define and understand the life course, and the two most relevant to our research are middle childhood (ages 6-11 years) and adolescent childhood (ages 12-18 years). The developmental changes that children go through as they approach and progress through adolescence are traditionally discussed in terms of their physical changes, cognitive and social development, and independence. Our research aims to focus on the transition between middle childhood and adolescence to determine their mental models of online deception. Wisniewski *et al.* [88] explain that a tension emerges between parents' control and teens' need for self-regulation. This issue is important as teens start to manage their own cybersecurity, and it is essential for them to have a nuanced understanding of their vulnerability online. Hence in our study we will focus on the mental models of those in the 11-14 age range, as emerging adolescents.

Research is needed to identify minors' understanding of cybersecurity and also of their awareness of so-called 'dark patterns' used to deliberately deceive them, especially as they move towards unsupervised use of online platforms.

We commence by reviewing the literature on dark patterns, human cognition, mental models and deceptive technologies in Section 2. Then, Section 3 examines the research related to cybersecurity behaviours of young teenagers. Since our research focuses on mental models, Section 4 considers mechanisms for elicitation of mental models in young learners. Section 5 considers how interventions can be designed to re-align mental models that are lacking or incorrect. Section 6.1 enumerates the ethical considerations for this kind of research, and Section 6.2 suggests a design for eliciting young learners' mental models related to online deception. Section 7 concludes.

2 Cybersecurity, Mental Models and Online Deception

2.1 Dark Patterns and Online Deception

Products that use dark patterns nudge people to make decisions that are not aligned with their own best interests. Some deceive users while others covertly manipulate or coerce them into choices that are not in their own best interests [50]. Dark patterns can compromise legal requirements, such as consent and privacy-by-design and legal principles, such as fairness and transparency [42].

Harry Brignull coined the term 'dark patterns' in 2010 and has published a website presenting some examples of actual dark

patterns at their hall of shame[1]. Others have also provided lists of dark patterns, including Shopify[2], Mathur *et al.* [48], and Luguri & Strahilevitz [46]. Here, we present Luguri and Strahilevitz's [46] category list:

Nagging: Repeated requests to do something the requesting party wants you to do for their benefit [44].

Social Proof: Using possibly spurious social norm influences.

1. *Activity messages:* False/misleading notice that others are purchasing, contributing [48].
2. *Testimonials:* False/misleading positive statements from customers [48].

Obstruction: Creating friction for the user in achieving their own goals.

1. *Roach Motel:* You get into a situation very easily, but then you find it is hard to get out of it (e.g., a premium subscription) [48, 10].
2. *Price Comparison Prevention:* The retailer makes it hard for you to compare the price of an item with another item, so you cannot make an informed decision.
3. *Intermediate currency:* Purchases in virtual currency to obscure cost [10].
4. *Immortal Accounts:* Account and consumer info cannot be deleted [8].

Sneaking: Engaging in some form of trickery.

1. *Sneak into Basket:* You attempt to purchase something, but somewhere in the purchasing journey the site sneaks an additional item into your basket, often through the use of an opt-out radio button or checkbox on a prior page [29, 48].
2. *Forced Continuity:* When your free trial with a service comes to an end but your credit card silently starts getting charged without any warning. In some cases this is made even worse by making it difficult to cancel the membership [29, 48].
3. *Hidden Costs:* You get to the last step of the checkout process, only to discover some unexpected charges have appeared, e.g., delivery charges, tax [29, 48].
4. *Bait and Switch:* You set out to do one thing, but a different, undesirable thing happens instead [29].
5. *Sneaking:* Attempting to misrepresent user actions, or delay information that if made available to users, they would likely object [48].

Interface Interference: Manipulating the interface to 'sludge' the user.

1. *Disguised Ads:* These are designed to look like regular content or navigation, in order to get users to click on them.
2. *Confirm shaming:* The act of making the user feel guilty to have them agree into opting into something. The option to decline is worded in such a way as to shame the user into compliance [10].

[1] `www.darkpatterns.org`, `https://www.darkpatterns.org/hall-of-shame`
[2] `https://www.shopify.com/partners/blog/dark-patterns`

3. *Urgency:* Imposing a deadline on a sale or deal, thereby accelerating user decision-making and purchases [48].
4. *Scarcity:* Signalling that a product is likely to become unavailable, thereby increasing its desirability to users [48].
5. *Trick questions:* While filling a form, the user is tricked into giving an answer they didn't intend. When glanced upon quickly the question appears to ask one thing, but when read carefully it asks another thing entirely [29, 48].

Forced Action: Forcing the user to do something tangential in order to complete their task.

1. *Friend Spam:* The product asks for your email or social media permissions under the pretence it will be used for a desirable outcome (e.g., finding friends), but then spams all your contacts in a message that claims to be from you.
2. *Privacy Zuckering:* You are tricked into publicly sharing more information about yourself than you really intended to. Named after Facebook CEO Mark Zuckerberg.
3. *Misdirection:* The design purposefully focuses your attention on one thing in order to distract your attention from another.

As observed from the list, dark patterns live at the intersection of three areas [50]: deceptive techniques [13], nudges [76] and social engineering [31, 34] (referred to by Narayanan *et al.* [50] as 'hacking'). The intersection is shown in Figure 1.

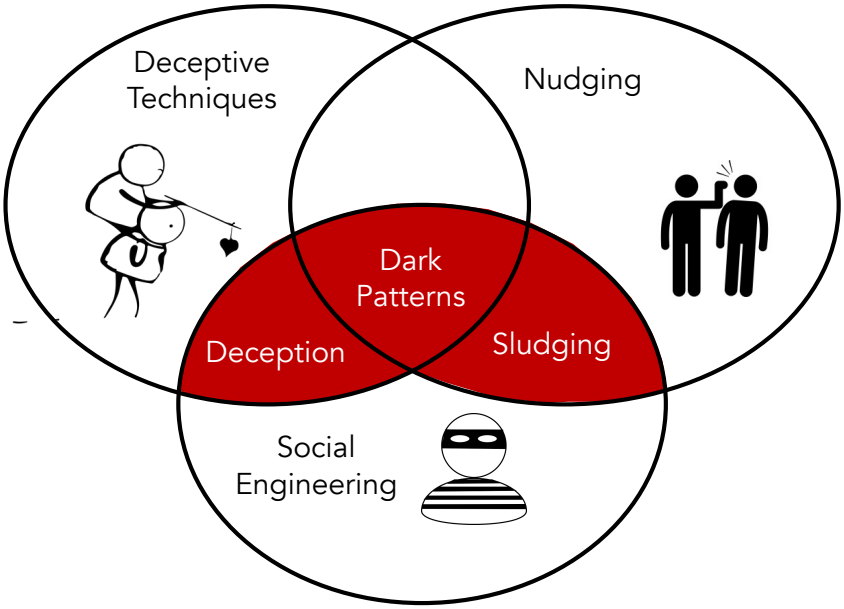


Figure 1: Situating Dark Patterns

To determine the power of dark patterns, Luguri and Strahilevitz [46] carried out an experiment with mild and aggressive dark patterns. Both patterns were embedded in a user interface that tries to persuade people to purchase an insurance policy against identity theft. An example of a mild dark pattern is to require people to give reasons for declining to 'guilt' them into buying the policy: e.g., "*Even though 16.7 million Americans were victimized by identity theft last year, I do not believe it could happen to me or my family*". An example from [46] is confirmshaming. An example of an aggressive dark pattern is that when they decline to purchase the policy, they are then forced to read information about identity theft by blocking them

from proceeding and showing a countdown timer while they read the text. Mild dark patterns were shown to be effective, but aggressive dark patterns were almost four times more effective than the usual user interface. The researchers report that whereas aggressive dark patterns generated a powerful backlash, mild dark patterns did not – suggesting that they are perhaps more effective in deceiving online users. Yet both of these demonstrate the efficacy of sludge in the hands of the unprincipled and unscrupulous.

2.1.1 AI-Powered Deceptive Technology

Dark patterns, as categorised above, and other forms of online deception are getting more advanced as technology changes and develops. One example is AI-supported voice synthesis technologies (e.g., Lyrebird [45]), which can imitate someone's voice or even change words in the original speech. This technology can be used for malicious purposes by perhaps bypassing biometric security processes [43], or sending malicious voice commands to Voice Operated Assistants, asking them to perform tasks and potentially jeopardising the security of the user's physical or online environment.

AI-based technologies may also lead to the weaponisation of social media platforms for facilitating spear-phishing efforts, selecting targets, and generating tailored machine-generated social media messages to these targets [41, 43]. Social media has particular characteristics that may increase a users' vulnerability to deception. Social media messages tend to be relatively brief and informally written. Studies have shown that bots on social media can be highly successful in deceiving online users and even, to an extent, security researchers [21]. There is evidence that oversharing information and engaging with strangers online is common in social media, both in teenagers and other age groups (e.g., [89]).

2.2 Human Cognition

Tversky and Kahneman [80] refer to two subsystems for human thinking: systems 1 and 2. System 1 is fast, intuitive and unconscious. System 2 is more calculating, marshalls evidence and considers all available options. System one is likely to produce more risky behaviours because it can produce more impulsive behaviours. The use of heuristics rather than a rational decision-making process has been cited as a factor explaining the privacy paradox, where stated privacy concerns do not convert to privacy-protective behaviours [1, 2]. This paradox has also been attributed to heuristic usage and biases [39].

Risk perception during interaction is subject to several potential influences [14]. First, the strength of user reaction to a security prompt is likely to be stronger where they are able to imagine the nature of the threat. For example, a prompt to create a stronger password will have a greater influence if the user is able to imagine the threat caused by using a more hackable password. Familiarity and previous stories they can identify with may influence how seriously a user takes security on a particular site or platform. Finally, accumulated malware-free experience may bias perceptions and reduce urgency related to staying attuned to possible threats and maintaining secure settings.

2.3 Mental Models

Any child or adult that navigates the online world makes security-related decisions. Internal mental models of online risks and threats drive the security decisions they make. Mental models are defined as: *"A concentrated, personally constructed, internal conception, of external phenomena (historical, existing or projected), or experience, that affects how a person acts"* [69, p. 16].

Cybersecurity mental models are composed of structures generated by users to understand the nature of cyber threats and their relationships to precautions for staying safe and secure online [85]. Everyone constructs mental models based on their lay understanding of concepts related to how the world works. Mental models influence user behaviour for both good and ill [9]. A survey showed that numerous mental models of key concepts are generated by lay users [85]. Previous research into mental models helped users to understand key concepts, such as 'viruses', where users are prompted to recruit knowledge from better understood domains [14].

Whilst mental models tend to partially map a familiar base domain to a target cybersecurity concept, it is argued that even 'wrong' mental models may be useful if they lead to secure behaviours [85]. Understanding how mental models are formed and associated with particular concepts can help designers to be aware of them in designing user interfaces.

Work on the provenance of user perceptions suggests that home computer users develop their mental models based on stories from friends and colleagues [85] or media stories. There is a tendency for users to focus on more newsworthy risks than other equally important risks [14]. This tendency is a manifestation of the availability heuristic [23]. More recently, research has also cited the impact of entertainment media on the formation of inaccurate or incorrect mental models [24]. Kang *et al.* [37] also found that experts had more elaborate and accurate mental models of online privacy threats than laypersons, but not everyone can be an expert.

3 Research into Young Learners & Cybersecurity

3.1 Education

Education in the UK is divided into Key Stages, with children in KS1 aged 5-7 years, KS2 aged 7-11 years, KS3 aged 11-14 and KS4 aged 14-16 years [28]. At each stage, there are specific skills, abilities and levels of knowledge that children should achieve for each subject. KS3 is of great interest to those researching the interdisciplinary areas of Human-Computer Interaction (HCI) and Cyberpsychology, as within this key stage, children transition to being legally able to have an online identity. On many social media platforms (such as Facebook), the legal age of use is 13 years; although in reality, many children are using accounts before this age [56].

3.2 Research into Teen Perceptions & Behaviours

There is limited research exploring children's understanding and mental models of these topics, as studies are often constrained by very strict ethical requirements, which challenges researchers [73]. Ethical restrictions have led to the majority of research focusing on teens 16+ years of age, or studies where parents or teachers have provided data relating to the children under 16

years of age that they teach or support [65].

The limited research that has been conducted on younger children tends to use small samples or a narrow age range, e.g., one year group. Very little research focuses on children between 11-14 years of age, which is a key transition period when children traditionally progress from primary to secondary school. Much of the research uses surveys that have primarily been designed by adults and appropriate questions designed for adults. It is crucial in any research involving children to use age-appropriate language to engage effectively with them.

3.3 Privacy Practices

Haber [30] highlights the risks of voice synthesis technology when smart toys and smart TVs also gather voice recordings, which may have a direct impact on children's online privacy and security. Related to social development, the desire for adolescents to develop new relationships and identity can lead to lapses in privacy as personal details are more easily divulged. For example, Youn [90] surveyed high school students and found their willingness to disclose personal information correlated with the perceived benefits of sharing.

Younger users are subject to pressures such as the need to reinforce self-esteem through social interaction [82], fear of missing out [12], or motivated by curiosity and exploration, treating risk as a learning experience [27, 87]. This suggests that privacy considerations were perhaps not uppermost in their minds and sludge could exploit this focus to extract more information than the teen ought to be divulging.

Privacy practices vary at different ages; younger children worry less about protecting their personal information and passwords than older children [68]. Younger children take more risks and are less aware of risks, but research examining older children shows that although they are more aware of the risks, they still engage in risky behaviour [58]. Hargittai and Marwick [33] found that while older children were more likely to engage in protective behaviour (for example, controlling who has access to their profile and providing false identity information). Despite this, they considered that they had limited control over their own data. Hargittai and Marwick discuss the so-called 'privacy paradox', whereby people claim that they care about their privacy, but their behaviour shows they divulge a lot of private details via social media. Teens might be also do this.

3.4 Home and Peer Influence

Teen security behaviours can be expected to be hugely influenced by the security behaviours at home. A 2007 study [26], aiming to assess the security perceptions of personal Internet users, asked 415 UK home users whether they had any children that used their home PCs, which are often viewed as a family resource. The survey revealed that 17% had systems that were used by children under 12, whereas 18% had computers that were used by children aged 12–18 (the overall proportion of respondents with children was 30%). These numbers are only expected to grow as home computer use becomes more accepted and increases with the pandemic forcing most school-related work to go online.

A particular vulnerability in teenagers may be due to a combination of unhelpful or poorly-formed mental models of risk and key security concepts and competing influences that may weaken their attention to security considerations. For example,

teenagers see digital environments as a place to share their accomplishments and spend time with peers and those with similar interests [5], confirming that interaction with peers and peer experience is a key influence on teenagers mental models and online behaviour. Security is likely to be a much lower priority.

3.5 Cybersecurity Practices

Dourish *et al.* [18] studied security in the wild to understand user attitudes and strategies for managing their security everyday. They observed that younger subjects, who have relatively more familiarity with computer systems (e.g., childhood exposure), express more confidence with their computer abilities. They also qualified as more pragmatic users, aware of the trade-off "obstacles" created through security measures in addition to their protection. Furthermore, beliefs in futility was also prevalent, assuming hackers will always be a step ahead. The way people manage their security takes the form of delegation, which includes delegation to a software, trusted person, organisation or institution. Other actions people took included security through obscurity (e.g., hiding the meaning of the messages in e-mails), or switching to more ephemeral mediums for sensitive communication (e.g, teenagers switching to phone instead of instant messaging), and keeping multiple online identities.

In terms of physical abilities, inputting passwords using a keyboard or touch screens becomes significantly easier as children become older; younger children might struggle to authenticate due to immature literacy. Relevant to their cognitive development, limited memory and spelling mean that younger children will use less complex passwords than older children [66]. However, because of the more simple passwords, younger children find it easier to remember passwords compared to older children [77].

Choong *et al.* [15] reviewed published research conducted since 2000 related to children and passwords. They also carried out a study with American children to uncover understanding of password principles. They reported that many of the children reported being confused about the need for passwords, and interestingly many referred to privacy and safety being guaranteed by a password rather than security. Similarly, Theofanos *et al.* [77] surveyed three distinct age-related groups of school children in the USA and discovered that many thought passwords would protect and keep them safe online, thereby conflating safety and security. Prior and Renaud [62] developed "best practice" ontologies to be used by early educators and parents and identified age-appropriate good practice principles for password creation.

Nicholson *et al.* [54] discovered that while teens had a working knowledge of cybersecurity theory, such as passwords, they did not necessarily convert their knowledge to actual behaviours. What is needed now is an understanding of how teens perceive the risks of their password management and other cyber security and privacy related practices as they progress through adolescence.

3.6 Summary

In summary, there is limited research exploring perceptions of cybersecurity from teens within the age of 11-14, and much of the existing research uses survey-based methods designed by adults. The majority of work focuses on password authentication. Therefore, in terms of scenarios indicated by this review, it would be useful to explore mental models of authentication for this age group, e.g., using non-survey methods to explore the gap between teens' password knowledge and their password behaviour.

Also, scenarios need to explore newer methods of authentication using biometrics (a search for perceptions of teens towards fingerprint and face recognition for authentication did not reveal any papers).

4 Eliciting Young Learners' Understanding of Online Sludge

An understanding of mental models can help us understand the inner thought processes of individuals. However, before we can hope to understand these mental models, we must first understand how to access them without changing them [84]. Extracting and understanding mental models is something which is known to be challenging [70], something considered even more difficult when minors are involved [47].

A variety of methods have been used to attempt to elicit mental models: asking someone to draw diagrams of their understanding of a particular topic can reveal their mental models, or they can be asked to arrange cards to demonstrate internal structures of knowledge [47]. Indirect elicitation, on the other hand, allows researchers to infer the individual's mental model from a questionnaire or interview data [47]. The most common form of elicitation involves the "teach-back individual interview" wherein participants explain or teach others as they carry out a drawing task [60]. However, we consider drawings the most appropriate medium for eliciting young learners' mental models, which is the topic of the next section.

4.1 Using Drawings to Elicit Mental Models

Drawing, as a method to elicit the mental models of children, is not a new concept. A wealth of literature exists relating to children's mental models of areas such as biology [72, 78, 79]. Vishkaie [83] argues that drawings are a viable way for children to express themselves, and drawings are mentioned specifically by Punch [64] in her discussion on ethical research with children. Driessnak [19] also talks about the power of listening to how children explain their drawings. Rowe and Cooke [71] tested four ways to measure mental models. One of these was diagramming, and it was found to be predictive of people's performance in a particular domain.

However, there are a number of considerations when asking people to take part in elicitation drawing tasks. For example, Prokop [63] demonstrate how the specificity of the instructions given to 10-14 year old children can greatly change the outcome of their drawings. Similarly, they outlined how the level of students' existing knowledge around a topic was strongly associated with the level of the drawing they were able to produce.

Far less research has been conducted with drawing mental models within the HCI sphere. A small but growing literature base has demonstrated the effectiveness of drawing for eliciting the mental models that children have in relation to technology, which we discuss next.

4.2 Mental Models of Technology

Kodama [38] invited 26 students aged between 10 and 14 to take part in a drawing activity in which they were to depict their understanding of how a Google search worked. They found that students had a poor understanding of Google and how it works, with interesting side findings such as students personifying Google. This poor understanding has possible subsequent

connotations for these students who may subsequently fall foul of misinformation, disinformation, or develop an over-reliance and unquestioning trust in the outcome of the search engine's results.

Pancratz and Diethelm [59] conducted a study wherein they invited 68 secondary school students to draw their conceptions of three computing systems (smartphones, video gaming consoles and robotic vacuum cleaners). Using a range of qualitative analytical methods, they identified misconceptions in the understanding of students, such as a lack of knowledge in how various components were connected. Interestingly they identified that the labelled components of their drawings were those the students most physically interacted with, such as buttons, displays, cameras and loudspeakers.

Denham [17] was one of the earlier researchers to use drawing methodologies to elicit technological mental models. Their rationale for doing so was that with younger students, a drawing task would be seen as more acceptable than interviews which might cause panic in students considering them to be test-like. Similarly, given the development of language in children, drawing may offer a stronger method of communication in which students can better express themselves without fear of "being wrong". An easy way to promote these drawings to a more detailed level is to ask the children to label their drawings as they draw; this assigns a clear meaning to the drawn aspects [59].

Brodsky [11] explored participants' mental models of the internet by collecting data in the form of pictures from adolescents (aged from 11–15 years) and young adults (aged from 18–22 years). Responses were categorised into four themes (technical components, functions, attributes, and feelings), and when these were compared for the two sample groups, they mostly did not differ by age, gender or social media use. The one area where they did differ was in terms of the models within the feelings category: the young adult participants' mental models more often cited negative feelings, such as online antisocial online behaviour and Internet addiction, compared to the adolescents. In both age groups, participants noted the ubiquity of the internet and [11] concludes by suggesting that further research could link these models of internet ubiquity in the lives of young people to further understand privacy and security risks.

4.3 Related Research

No existing research has sought to combine drawing methodologies with online sludge. Sludge is a relatively new area of research, evolving from the existing nudge literature [49]. Some recent research [50] has highlighted numerous forms of problematic sludge, known as "Dark Patterns", catalogued by individuals such as Harry Brignull (darkpatterns.org). Such catalogues demonstrate the numerous methods of online sludge, as described in Section 2.1. Although many examples of online sludge exist, very little is known about how children understand these and how this impacts whether or not they fall foul of such malicious content.

A study conducted by Oates *et al.* [55] analysed a repository of 366 drawings in response to the question 'what does privacy mean to you?', establishing emergent privacy themes and recurring symbols[3] Only a small number of these were from children aged between 11 and 13 (15) and a further 28 were in the age group 14-18. The analysis used five coding categories: metacodes, frameworks, visual symbols, metaphors and context. Metacodes encompassed attribute codes, identifying where the image provided was composite (with multiple sub-components or added text). Frameworks referred to identifiable mapping to established privacy frameworks, including Solove's Taxonomy of Privacy [75] and Westin's states of privacy [40]. Analysis from

[3]The repository is available at `https://cups.cs.cmu.edu/privacyillustrated/`.

drawings related privacy to digital and social media, with the frequency increasing steadily from 11 to 14 and beyond. .

Whilst there is a dearth of material focusing specifically on the 11-14 age group, a number of surveys and also participatory design exercises are reported involving subjects, some of whom were in the relevant age group. For example, a participatory design exercise with 14-17 year olds by Ashktorab and Vitak [6] used a variety of design techniques aimed at finding mitigation and prevention solutions against cyberbullying. These include five scenarios within which participants are presented with brief stories of cyberbullying on five different social media platforms. These included Facebook trolling, Snapchat 'flaming' where the individual is subject to repeated abusive messages, anonymous cyberstalking, exclusion, outing and trickery (e.g., posting a an intimate video within a school and on wider social media). They were then asked to create images of suggested solutions addressing the scenarios.

5 Cybersecurity Interventions to Encourage Self-Protective Behaviours

Cybersecurity interventions need to consider the beliefs about security and security behaviours, which have a significant effect on how users can be educated, and so, a more-is-better approach is expected to fail [86]. Especially for younger or less educated users, Wash *et al.* [86] argue that emphasising vulnerability or using scare tactics is unlikely to be helpful, as people in this demographic often do not believe there is anything they can do to change anything. Therefore, interventions intended to influence behaviours should focus on users whose beliefs are the weakest.

[86] particularly focuses on understanding awareness around viruses and hackers and works with Internet users that are above the age of 18. Still, their results are significant in showing that younger people and people with lower levels of education concern themselves more with direct and visible security threats. Interestingly, the majority of the participants held two distinct sets of beliefs: "hackers target home computer users" (84.5%), and "hackers target others" (71.3%). Self-protective behaviours included mostly using security software (67.4% at least often) and being careful on the Internet (69.1% at least often). However, expert security settings (e.g., updating software patches, or backing up data) are less popular (24.2% at least often). As observed in other studies, this study also confirms that believing that you can protect yourself on the Internet leads to more risky behaviour. Finally, users who feel psychological ownership for a computer are more likely to engage in self-protective measures.

Anderson *et al.* [4] study "conscientious cybercitizens, individuals who are motivated to take the necessary precautions under their direct control to secure their computer and Internet in a home setting." The authors studied the impact of several factors, e.g., the desire to protect one's own assets (e.g., computer), and desire not to cause harm to others, and peer pressure on security behaviours. Their study also shows that the greater the sense of ownership, the greater the desire to protect. Their work involving 18-24 years old participants shows that interventions with messages focused on the positive consequences of performing good security behaviour may be more persuasive than emphasising negative consequences.

Furman *et al.* [25], surveying participants to understand what they feel needs protection and how, find that people mostly protect personal and financial information and identities, including music, pictures, passwords, files, data, credit and debit card information, and Social Security numbers. Participants protected their home computers from viruses, hackers, the government, scammers, and colleagues (some participants used their home laptops for work). In addition, the study finds participants prefer

fingerprint-based authentication (53% for banking, 41% for shopping), compared to using username and passwords, hinting people opt for convenient and efficient methods of security for self-protection.

Dourish *et al.* [18] had similar findings, where user responses indicated an interest in efficient security solutions that handle multiple issues, rather than having to manage multiple solutions for a diversity of problems. In addition, their study recommends that security solutions be highly visible, emphasising functionality, and should be considered a collaborative accomplishment (rather than individual) as personal security is highly dependent on others' behaviours as well.

6 Researching Young Learners' Deception-Related Mental Models

6.1 Ethical Considerations for Research with Minors

"We will make certain that researchers and research itself are contributors to making 'a world fit for children'" [7, p. 19].

The ethical considerations for our research are as follows: **Harm:** Minors are classed as vulnerable research participants [64]. As such, we have to ensure that they are not harmed by any research they participate in.

Data: We have to pay special attention to the collection, use and retention of children's data [74]. In line with the European Union's GDPR, we have to embrace the principles of data minimisation. Moreover, we should anonymise the data at the earliest opportunity and store it on secure servers. Data can only be used for the purpose it was collected [57].

Age Appropriateness of Materials: The materials should be checked by a safeguarder as well as the local facilitator of the research to ensure that it is age-appropriate [57].

Informed Consent: It is essential to obtain informed and voluntary consent from parents [16]. The child participants should also freely give their assent to participating in the research study [22, 51].

Safeguarding: It is important to appoint a safeguarder to be present during research with children [35]. They will attend to children's responses to detect signs of distress such as falling silent or changing the subject [16]. If these are detected, the research can be terminated and the child reassured.

Payment and Compensation: The guiding principles of justice, benefit and respect underpin the need for research participants to be appropriately acknowledged, adequately recompensed and given fair returns for their involvement [20]. For example, a monetary payment might be inappropriate for young learners, with a participation certificate being better in acknowledging their contributions because actual monetary payment, directly given to the participant children or their parents, could potentially pressure, coerce or bribe them. The facilitating teachers, on the other hand, *should* be recompensed for their efforts.

6.2 Design for Activity to Reveal Young Learners' Deception-Related Mental Models

This section will use the nomenclature proposed by darkpatterns.org to refer to situations depicted by scenarios. We decided not to use the scenarios related to online shopping, given the age of our participants. We also included one scenario to reflect their understanding of a common warning - not a dark pattern but definitely a scenario they are likely to encounter.

The research questions we want to answer are:

RQ1: To what extent do young learners understand dark patterns used to deceive them online?

RQ2: What do young learners make of displayed warnings advising them not to visit a particular website?

RQ3: How effective are drawings in eliciting young learners' mental models of online deception?

RQ4: How effective is a discursive drawing production in improving young learners' mental models of online deception?

After this investigation reveals mental models, we plan to answer the fifth research question: **RQ5:** *What kinds of interventions are required to re-align mental models to reflect the reality of online deceptive techniques?*

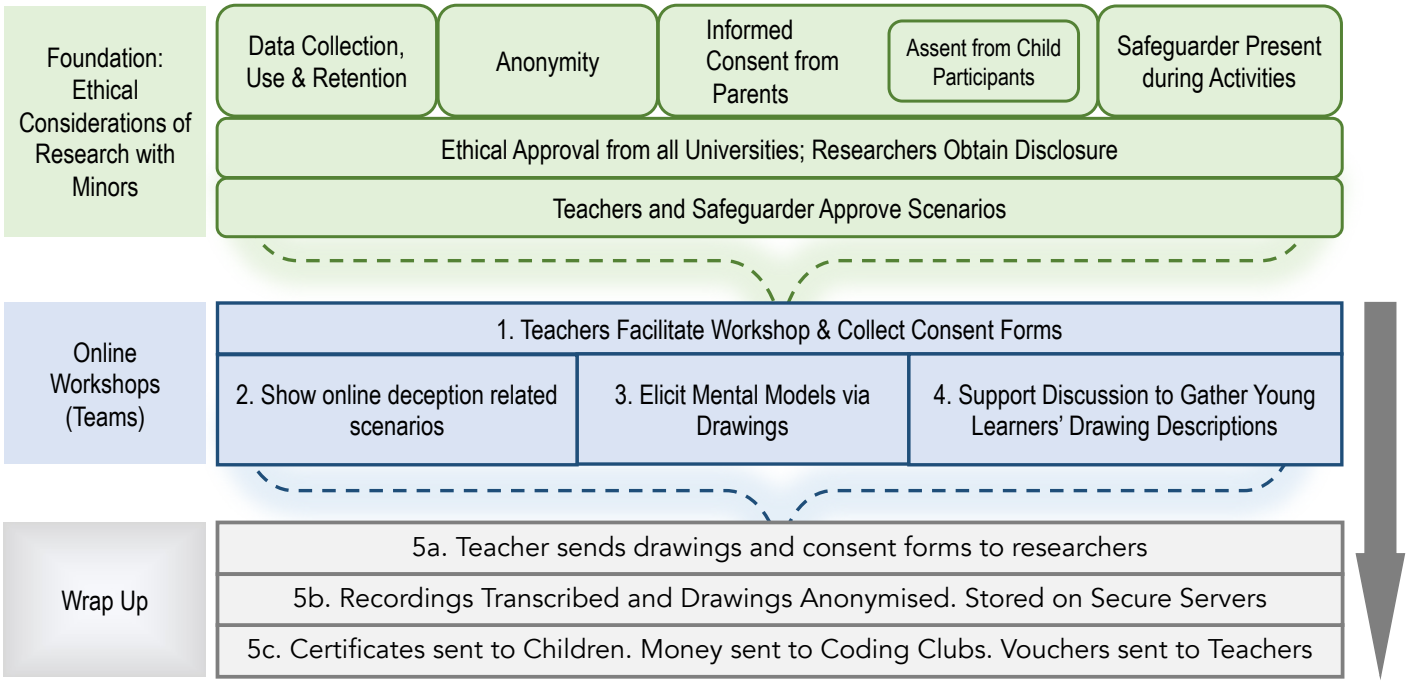


Figure 2: Research Design (Number refer to activities enumerated in Section 6.2)

Figure 2 presents the dimensions of the research, as informed by the discussion in the previous sections. In particular:

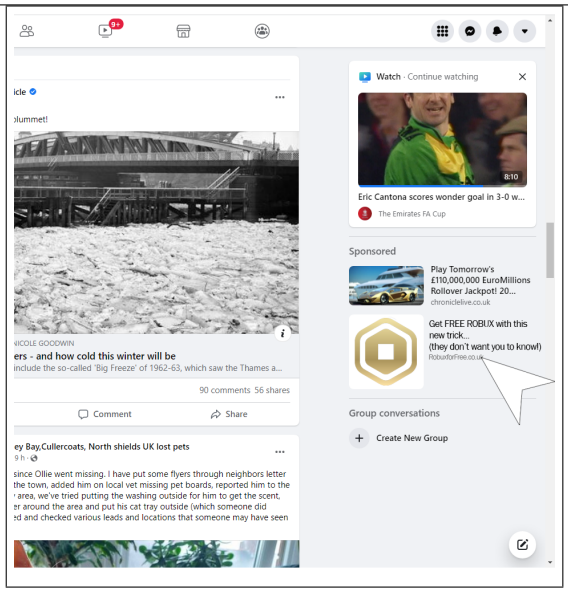
1. **Consent** – The facilitating teacher ensures that signed consent is obtained. This is obtained from parents and participating children assent to being involved. Teachers, too, sign consent forms.

2. **Online Sludge Scenarios** – The scenarios depicted in Table 1 will be shown to the young learners to reveal their mental models of online deceptive techniques.

3. **Activity 1** – Draw a picture related to what they think happens in the background if they click. They will also be asked to label their drawing to maximise researcher comprehension during analysis(Addressing RQ3)

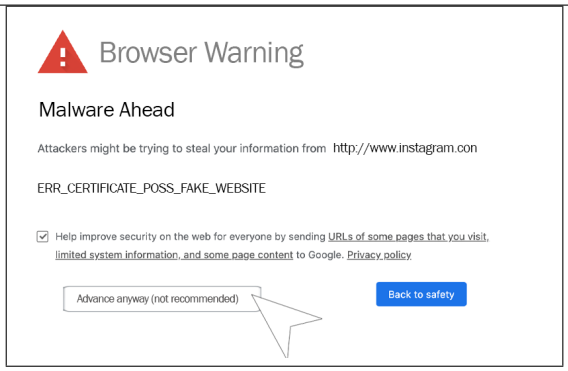
4. **Activity 2** – Young learners work together, discussing their drawings and producing a drawing together. This is because it is not enough merely to look at drawings, but also to hear what the children say about their drawings as well [22]. (Addressing RQ4)

Privacy Zuckering: You are tricked into publicly sharing more information about yourself than you really intended to. Named after Facebook CEO Mark Zuckerberg (retinal scan). (Addressing RQ1)



Bait and Switch: You set out to do one thing, but a different, undesirable thing happens instead. (Addressing RQ1)

Confirmshaming: The act of guilting the user into opting into something. The option to decline is worded in such a way as to shame the user into compliance. (Addressing RQ1)



Warning: to assess risk perceptions. (Addressing RQ2)

Bait and Switch: You set out to do one thing, but a different, undesirable thing happens instead. (Addressing RQ1)

Table 1: Dark patterns scenarios under study

5. Wrap Up – 5a. The teacher sends pictures of drawings to nominated researcher for anonymisation. The teacher sends the copies of all consent forms to the chief researcher. 5b. The schools receive monetary reward for participation, and the teachers receive vouchers in return for their facilitation of workshops. 5c. Young learners receive certificates of participation.

7 Conclusion

In this paper, we have reviewed the literature on dark patterns, and also the literature related to eliciting the mental models of young learners related to the online use of dark patterns. The main contribution is the design of an ethical informed workshop for revealing young learners' mental models of online 'sludge'.

Acknowledgement

This research was funded by a grant from SPRITE (`https://spritehub.org/2021/08/23/revealing-young-learners-mental-models-of-online-sludge/`).

References

[1] A. Acquisti, L. K. John, and G. Loewenstein. What is privacy worth? *The Journal of Legal Studies*, 42(2):249–274, 2013.

[2] Z. Aivazpour and V. S. C. Rao. Information disclosure and privacy paradox: The role of impulsivity. *SIGMIS Database*, 51(1):14–36, Jan 2020.

[3] J. Alemany, E. del Val, J. Alberola, and A. García-Fornes. Enhancing the privacy risk awareness of teenagers in online social networks through soft-paternalism mechanisms. *International Journal of Human-Computer Studies*, 129:27–40, 2019.

[4] C. L. Anderson and R. Agarwal. Practicing safe computing: A multimethod empirical examination of home computer user security behavioral intentions. *MIS Quarterly*, 34(3):613–643, 2010.

[5] M. Anderson and J. Jiang. Teens, social media & technology. *Pew Research Center*, 31(2018):1673–1689, 2018.

[6] Z. Ashktorab and J. Vitak. Designing cyberbullying mitigation and prevention solutions through participatory design with teenagers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 3895–3905, 2016.

[7] N. Bell. Ethics in child research: rights, reason and responsibilities. *Children's Geographies*, 6(1):7–20, 2008.

[8] C. Bösch, B. Erb, F. Kargl, H. Kopp, and S. Pfattheicher. Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns. *Proc. Priv. Enhancing Technol.*, 2016(4):237–254, 2016.

[9] G. L. Brase, E. Y. Vasserman, and W. Hsu. Do Different Mental Models Influence Cybersecurity Behavior? Evaluations via Statistical Reasoning Performance. *Frontiers in Psychology*, page 1929, 11 2017.

[10] H. Brignull. Types of dark pattern, 2020. `https://darkpatterns.org/types-of-dark-pattern.html`.

[11] J. E. Brodsky, A. K. Lodhi, K. L. Powers, F. C. Blumberg, and P. J. Brooks. "It's just everywhere now": Middle-school and college students' mental models of the Internet. *Human Behavior and Emerging Technologies*, 3:495–511, 10 2021.

[12] S. L. Buglass, J. F. Binder, L. R. Betts, and J. D. Underwood. Asymmetrical third-person effects on the perceptions of online risk and harm among adolescents and adults. *Behaviour & Information Technology*, 40(11):1090–1100, 2021.

[13] R. Calo. Digital market manipulation. *Geo. Wash. L. Rev.*, 82:995, 2013.

[14] L. J. Camp. Mental models of privacy and security. *IEEE Technology and Society Magazine*, 28:37–46, 2009.

[15] Y.-Y. Choong, M. Theofanos, K. Renaud, and S. Prior. Case study: exploring children's password knowledge and practices. In *Proceedings 2019 Workshop on Usable Security (USEC)*. Internet Society, Feb. 2019. Workshop in Usable Security and Privacy ; Conference date: 24-02-2019 Through 27-02-2019.

[16] V. E. Cree, H. Kay, and K. Tisdall. Research with children: sharing the dilemmas. *Child & Family Social Work*, 7(1):47–56, 2002.

[17] P. Denham. Nine- to fourteen-year-old children's conception of computers using drawings. *Behaviour and Information Technology*, 12:346–358, 1993.

[18] P. Dourish, R. E. Grinter, J. D. D. L. Flor, and M. Joseph. Security in the wild: User strategies for managing security as an everyday, practical problem. *Personal and Ubiquitous Computing*, 8:391–401, 2004.

[19] M. Driessnack. Children's drawings as facilitators of communication: a meta-analysis. *Journal of Pediatric Nursing*, 20(6):415–423, 2005.

[20] ERIC. Ethical guidance, 2021. `https://childethics.com/ethical-guidance/`.

[21] R. M. Everett, J. R. C. Nurse, and A. Erola. The anatomy of online deception: What makes automated text convincing? In *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, SAC '16, page 1115–1120, New York, NY, USA, 2016. Association for Computing Machinery.

[22] M. Fargas-Malet, D. McSherry, E. Larkin, and C. Robinson. Research with children: Methodological issues and innovative techniques. *Journal of Early Childhood Research*, 8(2):175–192, 2010.

[23] V. S. Folkes. The availability heuristic and perceived risk. *Journal of Consumer Research*, 15(1):13–23, 1988.

[24] K. R. Fulton, R. Gelles, A. McKay, Y. Abdi, R. Roberts, and M. L. Mazurek. The effect of entertainment media on mental models of computer security. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*, pages 79–95, Santa Clara, CA, Aug. 2019. USENIX Association.

[25] S. Furman, M. F. Theofanos, Y. Y. Choong, and B. Stanton. Basing cybersecurity training on user perceptions. *IEEE Security and Privacy*, 10:40–49, 3 2012.

[26] S. M. Furnell, P. Bryant, and A. D. Phippen. Assessing the security perceptions of personal internet users. *Computers and Security*, 26:410–417, 8 2007.

[27] M. Gardner and L. Steinberg. Peer influence on risk taking, risk preference, and risky decision making in adolescence and adulthood: an experimental study. *Developmental Psychology*, 41:625–635, 7 2005.

[28] GOV.UK. The national curriculum - gov.uk, 2021. `https://www.gov.uk/national-curriculum`.

[29] C. M. Gray, Y. Kou, B. Battles, J. Hoggatt, and A. L. Toombs. The dark (patterns) side of UX design. *Conference on Human Factors in Computing Systems - Proceedings*, 2018-April, 4 2018.

[30] E. Haber. The internet of children: Protecting children's privacy in a hyper-connected world. *U. Ill. L. Rev.*, page 1209, 2020.

[31] C. Hadnagy. *Social engineering: The art of human hacking*. John Wiley & Sons, 2010.

[32] Z. Hamdan, I. Obaid, A. Ali, H. Hussain, A. V. Rajan, and J. Ahamed. Protecting teenagers from potential internet security threats. In *2013 International Conference on Current Trends in Information Technology (CTIT)*, pages 143–152, 2013.

[33] E. Hargittai and A. Marwick. "What can I really do?" Explaining the privacy paradox with online apathy. *International Journal of Communication*, 10:21, 2016.

[34] HelpNet. Cybercriminals are manipulating reality to reshape the modern threat landscape - help net security. `https://www.helpnetsecurity.com/2021/08/05/cybercriminals-manipulating-reality/`.

[35] M. Hill, A. Lockyer, P. Morton, S. Batchelor, and J. Scott. Safeguarding children in scotland: the perspectives of children, parents and safeguarders. *Representing Children*, 13(3):169–183, 2002.

[36] R. K. Jones and A. E. Biddlecom. Is the internet filling the sexual health information gap for teens? An exploratory study. *Journal of Health Communication*, 16(2):112–123, 2011.

[37] R. Kang, L. Dabbish, N. Fruchter, and S. Kiesler. "My data just goes everywhere:" user mental models of the internet and implications for privacy and security. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, pages 39–52, 2015.

[38] C. Kodama, B. S. Jean, M. Subramaniam, and N. G. 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*, 20:403–432, 10 2017.

[39] S. Kokolakis. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, 64:122–134, 2017.

[40] P. Kumaraguru and L. F. Cranor. *Privacy indexes: a survey of Westin's studies*. Carnegie Mellon University, School of Computer Science, 2005.

[41] A. Laszka, Y. Vorobeychik, and X. Koutsoukos. Optimal personalized filtering against spear-phishing attacks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, 2015.

[42] M. Leiser. 'Dark Patterns': the case for regulatory pluralism. *Available at SSRN 3625637*, 2020.

[43] J. Li and N. Kaloudi. The ai-based cyber threat landscape. *ACM Computing Surveys*, 53:1–34, 2 2020.

[44] S. Livingstone and L. Haddon. EU Kids Online. *Zeitschrift Für Psychologie/Journal of Psychology*, 217(4):236, 2009.

[45] N. Lomas. Lyrebird is a voice mimic for the fake news era, Apr 2017. TechCrunch `https://techcrunch.com/2017/04/25/lyrebird-is-a-voice-mimic-for-the-fake-news-era`.

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

[47] A. M. Marhan, M. I. Micle, C. Popa, and G. Preda. A review of mental models research in child-computer interaction. *Procedia - Social and Behavioral Sciences*, 33:368–372, 2012.

[48] A. Mathur, G. Acar, M. J. Friedman, E. Lucherini, J. Mayer, M. Chetty, and A. Narayanan. Dark patterns at scale: Findings from a crawl of 11K shopping websites. In *Proceedings of the ACM on Human-Computer Interaction*, volume 3, pages 1–32. ACM New York, NY, USA, 2019.

[49] S. Mills. Nudge/sludge symmetry: on the relationship between nudge and sludge and the resulting ontological, normative and transparency implications. *Behavioural Public Policy*, pages 1–24, 12 2020.

[50] A. Narayanan, A. Mathur, M. Chetty, and M. Kshirsagar. Dark patterns: Past, present, and future. *Commun. ACM*, 63(9):42–47, Aug 2020.

[51] S. J. Neill. Research with children: a critical review of the guidelines. *Journal of Child Health Care*, 9(1):46–58, 2005.

[52] P. Newall, L. Walasek, E. Ludvig, and M. Rockloff. Nudge versus sludge in gambling warning labels. Technical report, PsyArXiv, 2020. `https://psyarxiv.com/gks2h/`.

[53] J. Nicholson, L. Coventry, and P. Briggs. Introducing the cybersurvival task: Assessing and addressing staff beliefs about effective cyber protection. In *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*, pages 443–457, Baltimore, MD, Aug. 2018. USENIX Association.

[54] J. Nicholson, J. Terry, H. Beckett, and P. Kumar. Understanding young people's experiences of cybersecurity. In *EuroUSEC*, 2021.

[55] M. Oates, Y. Ahmadullah, A. Marsh, C. Swoopes, S. Zhang, R. Balebako, and L. F. Cranor. Turtles, locks, and bathrooms: Understanding mental models of privacy through illustration. *Proceedings on Privacy Enhancing Technologies*, 2018:5–32, 10 2018.

[56] Ofcom. Children and parents: media use and attitudes report, 2020/21. `https://www.ofcom.org.uk/__data/assets/pdf_file/0025/217825/children-and-parents-media-use-and-attitudes-report-2020-21.pdf`.

[57] Ofsted. How we carry out ethical research with people, 2019. `https://www.gov.uk/government/publications/ofsteds-ethical-research-policy/how-we-carry-out-ethical-research-with-people`.

[58] U. Paluckaitė and K. Žardeckaitė-Matulaitienė. Adolescents' perception of risky behaviour on the internet. In *ICH&HPSY 2017 [electronic resource]: The European proceedings of social & behavioural sciences EpSBS: 3rd icH&Hpsy international conference on health and health psychology, July 5-7, 2017, Porto. London: Future Academy, 2017, vol. 30*, 2017.

[59] N. Pancratz and I. Diethelm. "Draw us how smartphones, video gaming consoles, and robotic vacuum cleaners look like from the inside": Students' conceptions of computing system architecture. *PervasiveHealth: Pervasive Computing Technologies for Healthcare*, 10 2020.

[60] G. Pask and B. C. Scott. Learning strategies and individual competence. *International Journal of Man-Machine Studies*, 4(3):217–253, 1972.

[61] M. Petticrew, N. Maani, L. Pettigrew, H. Rutter, and M. C. Van Schalkwyk. Dark nudges and sludge in big alcohol: behavioral economics, cognitive biases, and alcohol industry corporate social responsibility. *The Milbank Quarterly*, 98(4):1290–1328, 2020.

[62] S. Prior and K. Renaud. Age-appropriate password "best practice" ontologies for early educators and parents. *International Journal of Child-Computer Interaction*, 23-24:100169, 2020.

[63] P. Prokop, J. Fančovičová, and S. D. 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*, 4:75–93, 1 2009.

[64] S. Punch. Research with children: The same or different from research with adults? *Childhood*, 9(3):321–341, 2002.

[65] F. Quayyum, J. Bueie, D. S. Cruzes, L. Jaccheri, and J. C. T. Vidal. Understanding parents' perceptions of children's cybersecurity awareness in Norway, 2021. arXiv 2108.02512.

[66] J. C. Read and B. Cassidy. Designing textual password systems for children. In *Proceedings of the 11th International Conference on Interaction Design and Children*, IDC '12, page 200–203, New York, NY, USA, 2012. Association for Computing Machinery.

[67] V. Rideout and M. B. Robb. The common sense census: Media use by tweens and teens, 2019. Common Sense Media. `https://www.commonsensemedia.org/sites/default/files/uploads/research/2019-census-8-to-18-full-report-updated.pdf`.

[68] K. Rim and S. Choi. Analysis of password generation types in teenagers–focusing on the students of jeollanam-do. *International Journal of u-and e-Service, Science and Technology*, 8(9):371–380, 2015.

[69] L. Rook. Mental models: A robust definition. *The learning organization*, 20(1):38–47, 2013. https://doi.org/10.1108/09696471311288519.

[70] W. B. Rouse and N. M. Morris. On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, 100(3):349, 1986.

[71] A. L. Rowe, N. J. Cooke, K. J. Neville, and C. W. Schacherer. Mental models of metal models: a comparison of mental model measurement techniques. *Proceedings of the Human Factors Society*, 2:1195–1199, 1992.

[72] E. Rybska, S. Tunnicliffe, and Z. Chyleńska. Young children's ideas about snail internal anatomy. *Journal of Baltic Science Education*, 13:828–838, 12 2014.

[73] M. Sanjari, F. Bahramnezhad, F. K. Fomani, M. Shoghi, and M. A. Cheraghi. Ethical challenges of researchers in qualitative studies: The necessity to develop a specific guideline. *Journal of Medical Ethics and History of Medicine*, 7, 2014.

[74] J. Scott. The internet and protection of children online: time for change. *Canadian Journal of Law and Technology*, 9(1 & 2), 2011.

[75] D. J. Solove. A taxonomy of privacy. *U. Pa. L. Rev.*, 154:477, 2005.

[76] R. H. Thaler and C. R. Sunstein. *Nudge: Improving decisions about health, wealth, and happiness*. HeinOnline, 2007.

[77] M. Theofanos, Y.-Y. Choong, and O. Murphy. 'passwords keep me safe' – understanding what children think about passwords. In *30th USENIX Security Symposium (USENIX Security 21)*, pages 19–35. USENIX Association, Aug. 2021.

[78] S. D. Tunnicliffe and M. J. Reiss. Building a model of the environment: how do children see animals? *Journal of Biological Education*, 33(3):142–148, 1999.

[79] S. D. Tunnicliffe and M. J. Reiss. Building a model of the environment: how do children see plants? *Journal of Biological Education*, 34(4):172–177, 2000.

[80] A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, 1974.

[81] UNCTAD. Sludge, nudge and other consumer trends, 2018. `https://unctad.org/news/sludge-nudge-and-other-consumer-trends`.

[82] P. M. Valkenburg, M. Koutamanis, and H. G. Vossen. The concurrent and longitudinal relationships between adolescents' use of social network sites and their social self-esteem. *Computers in Human Behavior*, 76:35–41, 2017.

[83] R. Vishkaie. Companion toys for children: using drawings to probe happiness. *Interactions*, 28(4):39–43, 2021.

[84] M. Volkamer and K. Renaud. Mental models–general introduction and review of their application to human-centred security. In *Number Theory and Cryptography*, pages 255–280. Springer, 2013.

[85] R. Wash. Folk models of home computer security. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, pages 1–16, 2010.

[86] R. Wash and E. Rader. Too much knowledge? security beliefs and protective behaviors among united states internet users. In *Symposioum on Usable Privacy and Security*, pages 309–325, 2015.

[87] P. Wisniewski. The privacy paradox of adolescent online safety: A matter of risk prevention or risk resilience? *IEEE Security Privacy*, 16(2):86–90, 2018.

[88] P. Wisniewski, A. K. Ghosh, H. Xu, M. B. Rosson, and J. M. Carroll. Parental control vs. teen self-regulation: Is there a middle ground for mobile online safety? In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 51–69, 2017.

[89] W. Xie, A. Fowler-Dawson, and A. Tvauri. Revealing the relationship between rational fatalism and the online privacy paradox. *Behaviour & Information Technology*, 38(7):742–759, 2019.

[90] S. Youn. Teenagers' perceptions of online privacy and coping behaviors: A risk–benefit appraisal approach. *Journal of Broadcasting & Electronic Media*, 49(1):86–110, 2005.