

‘We’re not that gullible!’ Revealing dark pattern mental models of 11-12 year-old Scottish children

Karen Renaud
Cigdem Sengul
Kovila Coopamootoo
Bryan Clift
Jacqui Taylor
Mark Springett
Ben Morrison

© Authors 2024. This is the author's version of the work. The definitive Version of Record will be published in ACM Transactions on Computer-Human Interaction, <http://dx.doi.org/10.1145/3660342>

Renaud, K., Sengul, C., Coopamootoo, K., Clift, B., Taylor, J., Springett, M., & Morrison, B. (2024) *‘We’re not that gullible!’* Revealing dark pattern mental models of 11-12 year-old Scottish children. *ACM Transactions on Computer-Human Interaction*. Advance online publication. DOI: <https://doi.org/10.1145/3660342>

‘We’re not that Gullible!’ Revealing Dark Pattern Mental Models of 11-12 Year-Old Scottish Children

KAREN RENAUD, University of Strathclyde, Scotland; University of South Africa, South Africa; Rhodes University, South Africa; Abertay University, United Kingdom

CIGDEM SENGUL, Brunel University, United Kingdom

KOVILA COOPAMOOTOO, Kings College London, United Kingdom

BRYAN CLIFT, North Carolina State University, United States of America

JACQUI TAYLOR, Bournemouth University, United Kingdom

MARK SPRINGETT, Middlesex University, United Kingdom

BEN MORRISON, Northumbria University, United Kingdom

Deceptive techniques known as dark patterns specifically target online users. Children are particularly vulnerable as they might lack the skills to recognise and resist these deceptive attempts. To be effective, interventions to forewarn and forearm should build on a comprehensive understanding of children’s existing mental models. To this end, we carried out a study with 11-12 year old Scottish children to reveal their mental models of dark patterns. They were acutely aware of online deception, referring to deployers as being ‘up to no good’. Yet, they were overly vigilant and construed worst-case outcomes, with even a benign warning triggering suspicion. We recommend that rather than focusing on specific instances of dark patterns in awareness raising, interventions should prioritise improving children’s understanding of the characteristics of, and the motivations behind, deceptive online techniques. By so doing, we can help them to develop a more robust defence against these deceptive practices.

CCS Concepts: • **Security and privacy** → **Human and societal aspects of security and privacy; Social aspects of security and privacy**;

For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission

1 INTRODUCTION

The UK COVID-19 pandemic lockdowns and school closures¹ increased children’s online hours [46], often without direct adult supervision [52], which means that children now increasingly operate as autonomous agents when online. At the same time, parents are concerned about online risks to their children. Many doubt that the benefits of their child being online outweigh the risks (55% in 2020, down from 65% in 2015) [56].

Parental concerns about children’s online activities are well-founded. Similar to the physical world, the unscrupulous, the immoral, and the dishonest, i.e., ‘bad actors’, operate in the online world to carry out their nefarious and exploitative activities, using deceptive techniques to manipulate online users [10, 34]. The Internet frees these actors from the

¹The restrictions imposed by governments around the world during the COVID-19 pandemic resulted in different levels of restriction of personal movement. Over time, these changed from interaction just within one’s own household to later levels which allowed two other people and then small groups of people to interact.

Authors’ addresses: Karen Renaud, University of Strathclyde, Scotland; and University of South Africa, South Africa; and Rhodes University, South Africa; and Abertay University, United Kingdom, karen.renaud@strath.ac.uk; Cigdem Sengul, Brunel University, United Kingdom, cigdem.sengul@brunel.ac.uk; Kovila Coopamootoo, Kings College London, United Kingdom, kovila.coopamootoo@kcl.ac.uk; Bryan Clift, North Carolina State University, United States of America, bcclift@ncsu.edu; Jacqui Taylor, Bournemouth University, United Kingdom, jtaylor@bournemouth.ac.uk; Mark Springett, Middlesex University, United Kingdom, m.springett@mdx.ac.uk; Ben Morrison, Northumbria University, United Kingdom, benjamin.a.morrison@northumbria.ac.uk.

53 constraints of time and space so that they can now target millions of users, including children, across the globe. One of
54 their favoured techniques is the so-called ‘dark pattern’, which Brignull defines as ‘tricks used in websites and apps that
55 make you do things that you didn’t mean to’ [10].
56

57 Dark patterns pervade the online world. Consider that Di Geronimo [21] found that 95% of popular free-to-use mobile
58 apps available on Google Play Store, spanning categories of photography, family, shopping, social, music and audio,
59 entertainment, and communication, contained one or more dark patterns. Kowalczyk *et al.* [39] detected pervasive
60 dark patterns in popular Internet of Things (IoT) home devices, such as speakers, doorbells and cameras and Nouwens
61 *et al.* [54] found that they are prevalent in online consent management controls.
62

63 Children, given their youth and immaturity, might not yet be adept at spotting and resisting dark patterns. Yet,
64 they should be able to do this, especially 11-12-year-old children approaching the threshold age for digital consent, a
65 context where dark patterns are particularly common [54]. Ensuring that children are able to detect online deception is
66 therefore pivotal for reducing their online vulnerability and thereby preventing harm.
67

68 Interventions to forewarn and forearm are most effective when they build on a comprehensive understanding of
69 children’s mental models, which encode existing internal thought processes and beliefs [16]. At present, 11-12-year-old
70 children’s mental models of deceptive online techniques are poorly researched and imperfectly understood [65]. As
71 such, we do not know how best to target interventions to ensure that children develop the ability to detect and resist
72 online deception. The study we report on in this paper is one of the first to elicit and report on 11-12-year-olds’ mental
73 models of online deception.
74

75 The contributions of this study are as follows:
76

- 77 (1) An ethically-informed mixed-methods approach for eliciting Scottish 11-12-year-old children’s mental models
78 related to dark patterns and, by implication, online deception. The approach was carefully designed to elicit,
79 but not alter, existing mental models.
80
- 81 (2) Insight into Scottish 11-12-year-old children’s mental models of online dark patterns and deception. The findings
82 reveal an awareness of bad actors and their techniques but a lack of ability to distinguish between dark patterns
83 and genuine warnings.
84
- 85 (3) Suggestions for future work informed by these insights to help 11-12-year-old children to develop more accurate
86 and nuanced mental models of dark patterns/deception.
87
88
89
90

91 This study’s findings can inform human-computer interaction (HCI) researchers and practitioners. Gray *et al.* [31]
92 explain that user experience designers “could easily become complicit in manipulative or unreasonably persuasive practices”
93 (p.1) and argue for the HCI field to have a debate about applied ethics informing practitioners’ design activities. Gray *et*
94 *al.* [32] also consider deception in the context of consent banner design, where they pervade [54]. They argue that
95 studying deception in the design of these banners offers opportunities for bringing legal and ethical considerations into
96 HCI scholarship.
97

98 The rest of the paper is organised as follows. Section 2 reviews related research in this area. Section 3 outlines the
99 methodology of the study we carried out to reveal the mental models of 11-12-year-old Scottish children, as well as
100 the way we ensured ethical practice during the workshops. Section 4 reports on our results. Section 5 discusses the
101 findings, and Section 6 concludes.
102
103
104

2 RELATED RESEARCH

Covid-19 lockdowns drove a sharp increase in home technology usage by children [30]. Technology was used for leisure and also for structured activities such as remote schooling [43]. Ofcom reports that 71% of UK 8-11-year-olds and 94% of 12-15-year-olds use a smartphone to access the internet [58]. Based on the same report, 58% of children aged 3-15 use social media such as YouTube, TikTok, and Snapchat. Children are also likely to be particularly vulnerable to maliciously-targeted adverts and privacy risks from 'smart' toys [19].

In this section, we explore related research about legal approaches to the protection of children (Section 2.1). We then explain the nudge concept, its exploitation as 'dark patterns', and research into dark patterns exploiting children in Section 2.2. Section 2.4 concludes with a discussion of the nature and measurement of mental models with particular application to children.

2.1 Legal Approaches to Children's Online Protection

It is important to acknowledge the ongoing efforts to create a safer Internet for children and adults alike. In the UK, the controversial Online Safety Bill aims to make social media companies legally responsible for keeping children and young people safe online by making the risks and dangers posed to children on social media platforms more transparent and putting in measures to prevent children from accessing harmful and age-inappropriate content. The UK Council for Child Internet Safety (UKCCIS), a group of more than 200 organisations across government, industry, law, academia, and charity sectors, publishes regular guidance to help keep children safe online. The UK's Information Commissioner released 'Children's code design guidance' that sets out how online services that will be accessed by children should protect them online [35] while UNICEF released a report on ethical AI for children [84]. Finally, the UK government recently proposed the Online Safety Act 2023², an Act of Parliament to control online speech and media. The Act creates a duty of care for online platforms, requiring them to take action against potentially harmful content. In October 2023, the bill received Royal Assent [36], but we do not yet know how it will inform online child protection efforts.

2.2 Nudge

The concept of a choice architecture first emerged in Richard Thaler and Cass Sunstein's 2008 book, "Nudge: Improving Decisions about Health, Wealth, and Happiness", which explains all dimensions of the physical micro-environment within which a decisions are made [48]. Thaler and Sunstein [79] also introduced the concept of the 'nudge', a deliberate manipulation of this choice architecture designed to gently coax people to make wiser decisions. By definition, a nudge has to benefit the nudgee and have their implicit agreement to be nudged. Nudge examples include displaying the most secure WiFi at the top of the list on a Smartphone [82] or using visualisation to encourage stronger passwords [69]. The use of secure WiFi will prevent eavesdropping, and stronger passwords protect personal accounts. Both provide clear benefits to the nudgee. However, although nudges can benefit users, the technique can also be used for nefarious purposes when these principles are not respected [18, 45]. In these cases, they are termed 'dark' or 'deceptive' patterns, which are explained next.

2.3 Misuse of Nudging to Deceive

Harry Brignull originally coined the term 'dark patterns' in 2010 [10]. In 2023, he published a book where he argued for the use of the alternative term 'deceptive patterns', citing a definition proposed by Lisa Blunt Rochester referring to

²<https://www.legislation.gov.uk/ukpga/2023/50/contents/enacted>

157 them as: “*intentionally deceptive user interfaces that trick people*”³ [11, p.1]. Whatever the terminology, these deceptive
 158 techniques seek to persuade people to take some action to their own detriment. We shall use the term ‘dark patterns’ to
 159 refer to the concept in this paper.
 160

161 Dark patterns can compromise legal requirements, especially when they persuade the online user to grant consent
 162 unwisely, to click on links, to divulge information, or to make ill-advised purchases. Kahneman [37] suggests a dual
 163 system of thinking: System 1 and System 2. System 1 thinking is automatic and fast requiring little effort. System 2
 164 thinking is effortful, slow, and deliberate. Because nudges often target System 1’s automatic processing [6, 83], they do
 165 not engage the user’s conscious attention, which might help them to spot deceptive attempts. This means that nudges
 166 are unaware of their presence and influence [5].
 167

168 A number of websites are dedicated to dark patterns, including Brignull’s *hall of shame* [10], Shopify’s dark patterns⁴,
 169 and Thomas Mildner’s Dark Pattern Cheatsheet [51]. Mathur *et al.* [50] reviewed 11K shopping websites and then
 170 developed a taxonomy of dark pattern features as well as the cognitive biases they exploit. None of the identified
 171 patterns benefit the *nudgee*, the core requirement of a genuine nudge [79]. On the contrary, they are deployed to benefit
 172 the *nudger*. We used Mathur’s taxonomy to create the dark pattern scenarios used in this study (see Table 1).
 173
 174

175 2.3.1 Experiments with Dark Patterns.

176 Dark patterns come in various forms, with some designed to coerce gently (e.g., Confirm Shaming [10, 51]) and others
 177 designed to force users into taking some action (e.g., Coercion [17]). The former is ‘mild’, the latter ‘aggressive’. Luguri
 178 and Strahilevitz [45] carried out an experiment using mild and aggressive dark patterns embedded in a user interface
 179 that sought to persuade participants to purchase an identity theft insurance policy. The mild shame-based dark pattern
 180 required people to choose pre-defined reasons for declining the policy: e.g., “*Even though 16.7 million Americans were*
 181 *victimized by identity theft last year, I do not believe it could happen to me or my family*” (p. 62) (‘Confirm Shaming’
 182 [10, 51]). The aggressive dark pattern forced users to read information about identity theft and prevented them from
 183 proceeding and showing a countdown timer while they read the text. Mild dark patterns were somewhat effective, with
 184 aggressive dark patterns being almost four times more effective at prompting desired behaviours. However, aggressive
 185 dark patterns generated a powerful backlash, unlike mild dark patterns. Importantly, both demonstrate the destructive
 186 power of these techniques in the hands of bad actors, with the mild pattern being more insidious because users are less
 187 likely to notice them or detect their influence.
 188
 189
 190
 191

192 2.3.2 Children and Dark Patterns.

193 Children are certainly being exposed to deception online. Research by Ofcom in 2021 [57] found that 37% of 12-year-old
 194 children reported having seen misleading news online or while on social media. 33% of the 12-year-olds were unsure,
 195 and only 29% did not think they had seen fake media, which does not reliably point to the absence thereof. Moreover, it
 196 has been reported that children experience difficulties identifying fake media. For example, Statista reported that when
 197 a sample of UK children were asked whether a news story on social media was true, 43% found it quite difficult, with 9%
 198 finding it very difficult [77]. During October 2023, a number of US States sued Meta over children’s mental health and
 199 privacy. They specifically mention dark patterns that are harmful to children’s well-being (such as ‘likes’ and haptic
 200 notifications) [85].
 201
 202
 203
 204

205 ³Rep. Lisa Blunt Rochester on DETOUR Act - [https://www.warner.senate.gov/public/index.cfm/2021/12/lawmakers-reintroduce-bipartisan-bicameral-](https://www.warner.senate.gov/public/index.cfm/2021/12/lawmakers-reintroduce-bipartisan-bicameral-legislation-to-ban-manipulative-dark-patterns)
 206 [legislation-to-ban-manipulative-dark-patterns](https://www.warner.senate.gov/public/index.cfm/2021/12/lawmakers-reintroduce-bipartisan-bicameral-legislation-to-ban-manipulative-dark-patterns)

207 ⁴<https://www.shopify.com/partners/blog/dark-patterns>
 208

209 Although researchers have gained insights into the specific risks to children in online settings [68], very little
210 literature has explored the extent to which children themselves understand specific threats, what their risk management
211 approaches are, or, indeed, details of their underlying mental models. The latter are likely to impact their coping
212 mechanisms when faced with online threats. A few studies demonstrate the positive potential of training children to
213 avoid threats such as phishing [41] and promising outcomes of security awareness training [2]. However, crucially, we
214 do not yet know whether children are aware of dark patterns and online deceptive attempts more generally.

215 Recent reviews, such as Vissenberg *et al.*'s [89] conclude negative online experiences impact young people's well-
216 being negatively but are also essential to developing cyber resilience. The systematic review by Livingstone *et al.*
217 [44] demonstrated that children with greater digital skills were more likely to be exposed to online risks. However,
218 establishing specific links to consequent harms⁵ proved to be challenging without dedicated follow-up research.
219 Nevertheless, this literature base is growing rapidly, with the impetus for understanding children's cybersecurity
220 knowledge increasing [26, 65]. Below, we discuss how mental model studies offer a possible route to exploring this
221 knowledge.
222
223
224
225

226 2.4 Mental Models

227 Mental models are defined as: “A concentrated, personally constructed, internal conception of external phenomena (historical,
228 existing or projected), or experience, that affects how a person acts” [72, p. 16]. In essence, mental models reflect our
229 understanding of topics and inform our choices [7] and online decision-making [90]. In the following, we discuss
230 cybersecurity mental models, which are composed of structures generated by non-expert users to understand complex
231 topics such as cyber threats [91].
232
233
234

235 2.4.1 Cyber-Related Mental Model Research.

236 Previous research has discovered that people develop their mental models based on stories from friends and col-
237 leagues [91] or media stories [13]. This has disadvantages. In the first place, there is a tendency for people to focus on
238 more newsworthy risks [14], a manifestation of the availability heuristic [28]. Coverage of online deception is likely
239 to be patchy and might focus on sensationalism rather than the more mundane yet prolific deceptive techniques. In
240 the second place, recent research has reported that this reliance on entertainment media has led to the formation of
241 inaccurate or incorrect mental models [29]. A good example is the James Bond movie *Skyfall* which shows an expert
242 connecting a suspect laptop to a secure intelligence network. If people consider this credible, this might lead to risky
243 behaviours.
244
245

246 Wash and Rader [92], in a study of adult home computer users, identified two broad partially overlapping groups
247 of folk models (non-expert theories): (1) virus models, and (2) hacker models. Within these groups, they identified
248 models referring to the perceived actors and motivations behind their behaviour. These included the ‘buggy’ model (due
249 to software flaws), ‘mischief’ (due to mischief-mongers), ‘crime’ (intended to obtain sensitive information), ‘burglar’
250 (stealing financial data), ‘vandal’ (causing damage for showing off) and ‘big fish’ (targeting rich or important individuals
251 for attacks). Mischief and vandal are seen as types of virus and hacker models, respectively, but are very similar in
252 that the motivation of the actor is seen as to show off rather than to commit an identity or financial crime. Wash and
253 Rader [92] argue that these folk models tend to be inaccurate but may (or may not) lead to positive security decisions.
254 For example, a questionable belief that hackers are young, immature people showing off to friends may nonetheless
255
256
257

258 ⁵Examples of online risks include identity theft, GPS tracking enabling one's whereabouts to be visible and for them to be potentially followed or located,
259 or accidentally downloading illegal content from an insecure file-sharing network [14].
260

261 lead to an appropriate level of caution. Conversely, a belief that hackers only ‘go after the big fish’ may lead people to
262 dangerous complacency about their own vulnerability. A more realistic understanding of the motivations of bad actors
263 would bring perceptions into line with reality.
264

265 2.4.2 *Eliciting Mental Models.*

266 Extracting and understanding people’s mental models is challenging [16, 73], even more so when children are in-
267 volved [47]. Moreover, the act of measuring mental models risks changing them [74, 90]. Hence, we need to consider
268 carefully how to access them non-invasively [90]. A variety of methods have been used to elicit mental models, e.g.,
269 asking someone to draw diagrams [55, 75] or to engage in a participatory design exercise [3]. Alternatively, participants
270 can be asked to arrange cards to reflect internal knowledge structures [47]. A promising technique for eliciting complex
271 mental models is to ask people to create a drawing of a complex topic. The most common form of elicitation involves
272 the ‘teach-back individual interview’ wherein participants explain or teach others as they carry out a drawing task [60].
273 This allows researchers to observe, minimising the risk of altering the participant’s existing mental model. The next
274 section reviews the use of drawings in this context.
275
276
277

278 2.4.3 *Using Drawings to Elicit Children’s Mental Models.*

279 Using drawings to elicit mental models is not a new idea [76, 80, 81], drawings being a viable and familiar way for
280 children to express themselves [88]. Driessnak’s [24] meta-analysis identifies drawing as a robust way of measuring
281 mental models. Denham [20] was one of the first to use drawing methodologies to elicit mental models, arguing that a
282 drawing task is less threatening to children. Moreover, it is also considered an ethical way of carrying out this kind of
283 research [64].
284
285

286 Several studies employed drawings to explore perceptions of electronic systems and digital landscapes. Pancratz and
287 Diethelm [59] used drawings to identify misconceptions related to the functioning of electronic systems. Kodama [38]
288 used drawings to elicit an understanding of Google search, revealing a poor understanding of the underlying mecha-
289 nisms, which is likely to lead to unquestioning belief in misinformation and disinformation returned in search results.
290 Brodsky [12] successfully used diagrams to compare the mental models of the internet of different age groups (11-15
291 years and 18-22 years). They conclude that drawings elicit rich data to support qualitative analysis. The participant
292 responses were categorised into four themes: (1) technical components, (2) functions, (3) attributes, and (4) feelings.
293 When these were compared for the two sample groups, they mostly did not differ except for the ‘feelings’ category:
294 the young adult participants’ mental models more often cited negative feelings, such as antisocial online behaviour
295 and Internet addiction, compared to the adolescents. Both age groups noted the ubiquity of the internet, and Brodsky
296 concludes by suggesting that further research could link these models of internet ubiquity in the lives of young people
297 to further understand privacy and security risks [12].
298
299

300 Although the literature demonstrates the effectiveness of drawing as a methodology for eliciting children’s mental
301 models, there are a number of considerations. One is to be aware of the danger of altering existing models. For example,
302 Prokop *et al.* [63] demonstrate how the specificity of the instructions can influence drawings, highlighting the need to
303 be aware of how the task is framed in the experimental design. Moreover, working with children in any capacity raises
304 the need for ethical rigour and consideration of safeguarding, and we planned all our activities with this in mind.
305
306
307

308 3 STUDY

309 Our aim was to reveal children’s mental models of dark patterns. As such, we showed them three deceptive patterns and
310 a genuine warning to reveal the depth of their mental models. The use of a combination of drawings and explanations
311
312

is suggested by Pask and Scott [60]. We thus gathered drawings, as well as transcripts of discussions of the drawings, during workshops, to support analysis of children's mental models of online deception.

The scenarios we used included dark patterns with consequences ranging from minor (watching an advert) to severe (loss of credentials or privacy), as well as a genuine warning. Due to ethical and pandemic constraints (the inability to be present in person), teachers facilitated the workshops while we listened via a Microsoft Teams call to provide guidance. Our study was designed to answer the following research questions:

RQ1: Dark Pattern Detection: Can 11-12 year-old children:

RQ1a: detect different dark patterns?

RQ1b: correctly distinguish a genuine warning from a dark pattern?

RQ2: Dark Pattern Actors: How well do 11-12 year-old children understand the motivations of the actors who are using dark patterns to deceive them?

RQ3: Dark Pattern Actions & Consequences: What actions are bad actors perceived to take, and what consequences do children imagine will occur if they are deceived by a dark pattern?

3.1 Scenarios

We chose the dark pattern scenarios with great care to deliver insights into the children's mental models. These are summarised in Table 1. The scenarios were reviewed by Education Scotland⁶ and approved by ethics review boards of all authors' institutions.

Scenario 1: 'Privacy Zuckering' – This dark pattern is mentioned by Brignull [10] and by Bösch *et al.* [6]. We wanted to include a scenario that is related to a privacy dark pattern because: (1) these kinds of privacy-invasive patterns are particularly insidious, (2) people can easily lose their privacy, and once lost, privacy cannot be retrieved. Our design was inspired by a real case of a cryptocurrency exchanging free currency for iris readings [42]. The idea was to see whether the children realised that their biometric (eye scan) ought to be preserved and the possible consequence of this information being leaked to other entities. This information leak has possible cybersecurity consequences if the eye biometric is used for authentication.

Scenario 2: 'Bait & Switch' – in this dark pattern, a deceptive link appears to lead the user to something desirable, while it actually sends them somewhere unpleasant. This pattern is also mentioned by both Brignull [10] and Greenburg *et al.* [33]. In this scenario, the children are offered free Robux⁷ but redirected to a fake website. A possible consequence of this action may be a drive-by download or loss of credentials if the fake website is convincing enough to elicit these. Hence, the cybersecurity consequences may be potentially severe.

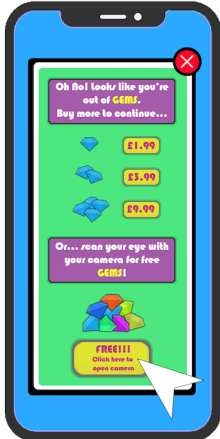
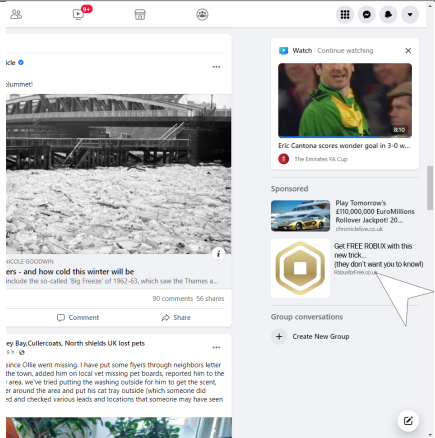
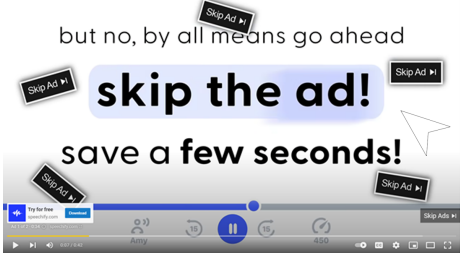
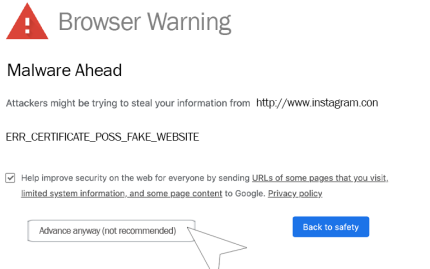
Scenario 3: 'Confirm Shaming' – which attempts to manipulate the viewer into diverting to the advertisement [10, 51]. This is a relatively mild dark pattern that guilt the user into opting for something. The option to decline is worded in such a way as to shame the user into acting to benefit the dark pattern deployer. We designed this scenario based on children's reported use of YouTube [23]. The scenario does not offer a particularly enticing bait and has no real cybersecurity consequences. In reality, they might see an advert or be persuaded to buy something, but their device and/or information will not be breached.

⁶National body for supporting quality and improvement of learning and teaching in Scottish education (<https://education.gov.scot/>).

⁷A very popular game amongst 10-12 years olds in the UK [22]

Table 1. Dark Pattern Scenarios under Study

365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416

Name	Description & Consequences	Scenario Image
<p>Scenario 1: Privacy Zuckering In terms of Mathur <i>et al.</i>'s [50] taxonomy, this one is: Asymmetric, exploiting the Framing Effect.</p>	<p>Description: The children are given the option to scan their eyes for free gems in the game. With this scenario, you can be tricked into publicly sharing more information about yourself than you really intended to.</p> <p>Consequences: Eye iris captured and possibly leaked to other 3rd parties. Possible cybersecurity consequences if the eye biometric is used for authentication.</p>	
<p>Scenario 2: Bait & Switch: In terms of Mathur <i>et al.</i>'s [50] taxonomy, this one is: Covert exploiting the Scarcity Effect.</p>	<p>Description: The children are offered free Robux but should expect to be redirected to a fake website. So, you set out to do one thing, but a different, undesirable thing happens instead.</p> <p>Consequences: Drive-by download or loss of credentials if the fake website is convincing enough to elicit these. Potentially severe cybersecurity consequences.</p>	
<p>Scenario 3: Confirm Shaming: In terms of Mathur <i>et al.</i>'s [50] taxonomy, this one is: Asymmetric exploiting the Bandwagon Effect.</p>	<p>Description: The children are shown a YouTube video with a message daring them to skip the ad. The option to decline is worded in such a way as to shame the user into acting to benefit the dark pattern deployer.</p> <p>Consequences: In this scenario, the children might see an advert and be persuaded to buy something. No cybersecurity consequences.</p>	
<p>Scenario 4: Genuine Browser Warning Warning No dark pattern</p>	<p>Description: A genuine browser warning to test for false positives.</p> <p>Consequences: If the warning is ignored, the risk is continuing to a fake website that the warning is related to. Potentially severe cybersecurity consequences.</p>	

417 **Scenario 4: ‘Genuine Browser Warning’** – included to see whether children would raise false positives by being
 418 suspicious of this warning instead of realising it was actually legitimate. A possible consequence of misclassification
 419 is visiting the fake website the warning is related to, and this may well trigger potentially severe cybersecurity
 420 consequences.
 421

422 3.2 School Recruitment

423 We recruited primary school classes from Scotland to participate in our workshops via educational authorities, who
 424 advertised the research study to all the schools in their districts. Two people from the educational authorities approved
 425 the scenarios we used during the workshops. We provided a short training session to participating teachers about the
 426 study and how we expected the activities to play out in the classroom. One school was in Aberdeenshire, two were in
 427 the Strathclyde area, and the others were in North Lanarkshire.
 428

429 All schools in Scotland follow the mandated Scottish curriculum: the *Curriculum for Excellence*, which includes
 430 ‘Digital Literacy.’ The guidance provided for 11-12-year-old children includes: “*I can keep myself safe and secure in online
 431 environments, and I am aware of the importance and consequences of doing this for myself and others*” [25]. While the
 432 curriculum is centrally provided, it is up to each school to decide how to teach these principles, which means that there
 433 will be some inevitable variability in terms of specific concepts taught.
 434

435 The recruited classrooms had a varied number of 11-12 year old children. Given that we were not present and did
 436 not count the number of students participating in each workshop, we used the number of drawings submitted per
 437 workshop as a proxy (Table 2). We carried out 7 workshops in 7 different classes.
 438

439 Table 2. Numbers of Participants in Each Workshop (WS)

WS1	WS2	WS3	WS4	WS5	WS6	WS7
14	8	32	28	30	16	24

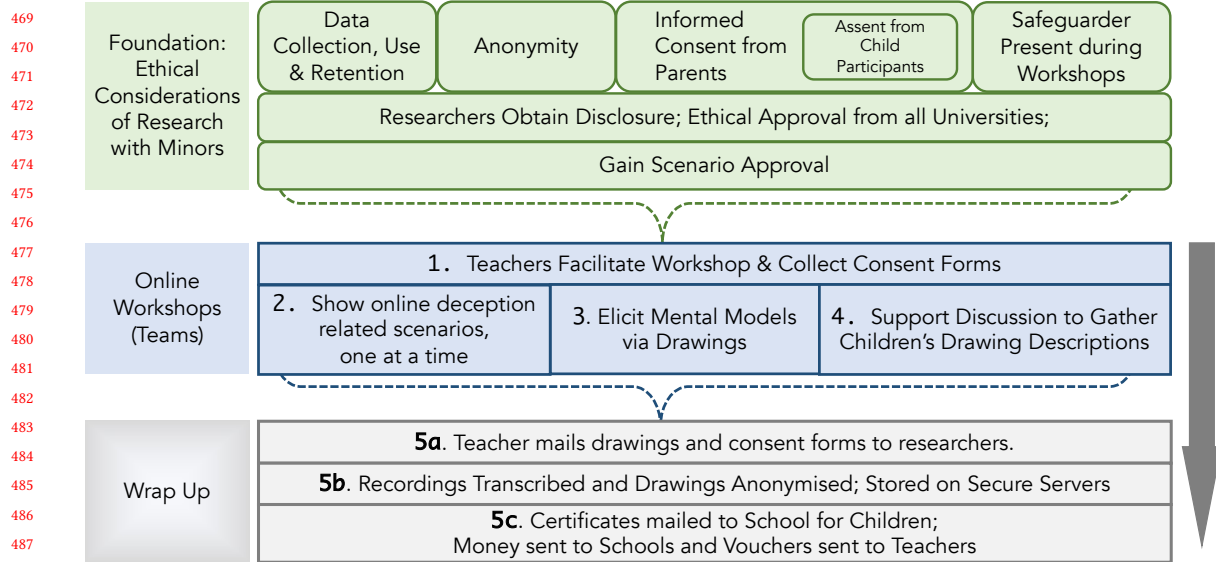
440 We note that we chose not to capture demographic attributes such as the gender and ethnicity of the participants
 441 because we considered it crucial to guarantee their anonymity. Future research should investigate gender and/or
 442 ethnicity’s impact on the ability to detect online deception.
 443

444 3.3 Procedure

445 We carefully designed the studies to the highest ethical standards, especially since we ourselves had to attend virtually
 446 and rely on teachers to facilitate the workshops on our behalf. As such, we developed a rigorous methodology for
 447 carrying out research remotely with classes of children. Ethical approval was gained from the ethical review boards of
 448 all participating institutions before we commenced. We provide a substantive discussion of ethical considerations and
 449 our methods for resolving them in Morrison *et al.* [53](summarised in Figure 1). A safeguarder, who had been vetted by
 450 the UK’s Disclosure and Barring Service, was present during all workshops to ensure that the children’s safety was
 451 monitored and assured.
 452

453 Figure 1 presents the dimensions of the research. In particular, for each workshop, we followed the protocol below:

- 454 **1. Consent.** The facilitating teacher ensured that signed consent was obtained from the parents of every par-
 455 ticipating child. The children themselves assented to participating. Children were informed that they could
 456 withdraw at any time without giving a reason. Teachers, too, signed consent forms.
 457
 458



489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520

Fig. 1. Research Design

All participating schools and teacher facilitators received a gift voucher, and every child received a certificate of participation. The certificates were mailed to the school, where the teacher added the children's names. One school received a special certificate for the entire class for being the best class to participate in the research project.

2. **Teacher Facilitation.** At each workshop, the teachers facilitated the workshops in the classroom, while two to three researchers, in addition to the safeguarder, joined the Microsoft Teams meeting used for coordination and guiding teachers. The Microsoft Teams meeting used only microphones and no cameras, and therefore, the researchers could not observe the classroom and relied on the teachers to facilitate the activities detailed below.
3. **Scenarios and Structured Drawing Activity.** The teacher showed the scenarios depicted in Table 1 to the children, one at a time. The children were asked to draw what they thought would happen in three steps if they clicked on the area highlighted by an arrow shown in the scenarios. All children participated in this activity.
4. **Unstructured Interview.** The child volunteers participated in an unstructured audio interview with the remote researcher to describe and discuss their drawings and interpretations of the scenarios [27]. This group was composed of volunteers and children nominated by the facilitating teachers. The researchers did not have any influence on who was chosen to engage with them.
5. **Wrap Up.** 5a. The teacher mailed drawings to the nominated researcher for anonymisation and consent forms to the chief researcher. 5b. The schools received a monetary reward for participation, and the teachers received vouchers in return for their facilitation of workshops. 5c. Chief researcher mailed blank participation certificates to schools for teachers to add names and issue to the children.

Due to the remote nature of the workshops and the lack of video footage, the study had several limitations, which we detail in Section 5.2.

3.4 Analysis

Our data collection methodology resulted in two types of data: (1) children's drawings, and (2) transcripts of the audio data collected during the workshops when children spoke about their drawings and answered questions from the teachers or the researchers. As outlined in the previous section, we were unable to match the children's drawings to their verbal explanations of their drawings. We thus analysed transcripts and drawings separately before discussing the entire study's findings and implications in Section 5. The analyses are detailed in the next subsections. We had to exclude the first workshop from the drawing analysis due to a number of drawings being missing from the pack sent to the researchers.

3.4.1 Drawing Analysis Method.

Framework Analysis (FA) was employed to analyse the drawings. FA is a matrix-based analytical framework that provides consistency and transparency across a data set [71]. This analytical approach draws inspiration from and aligns with Kodama *et al.*'s [38] analysis of drawings of mental models. In FA, a framework is developed that classifies and organises data into key themes, concepts, or categories. Ritchie and Spencer [70] suggest that FA is useful for analysing data derived from research questions that are contextual (e.g., examining the form and nature of knowledge and experience) and diagnostic (e.g., exploring reasons or influences upon knowledge and experience). Our FA approach consisted of 5 phases.

Phase 1: familiarisation. We began our analysis with a more inductive, open description of each drawing, a transcription of the visuals that strove to capture illustrations and how they were represented in textual form. Two groups of two researchers wrote descriptions of the drawings as the data to be analysed; within those groups, each researcher independently read data for initial description, transcription, and initial impressions of a drawing, which provided a visual and textual basis for making sense of a drawing, and then discussed potential themes (see Table 3).

Phase 2: framework construction. From these initial readings and conversations, we developed a framework through which the data were re-read. The purpose of this was to organise the data in a consistent, manageable, and meaningful way across all drawings. This enables speedier retrieval, exploration, and analysis during later stages. This phase produced four analytical themes, which were subsequently used to inductively code the descriptions of children's drawings in Phase 3. The themes revealed children's mental models of (1) a tit-for-tat exchange in mirroring the scenario (effectively replicating what was in the scenario image), (2) imagining beyond what was presented in the scenario with bad actor intentions, (3) identification of potential account compromise, and (4) leaked sensitive and personally identifiable information, as summarised in Table 3. Potential consequences of the scenario (such as hacking) were also identified as another theme.

Phase 3: indexing and sorting. With the framework established, the two groups of researchers re-read the drawings and transcripts of the drawings against the framework, organising the data into the framework categories. The research team systematically applied this across all drawings. Each coding pair was randomly allocated three workshops to code. Then each individual within a pair coded the data independently, and then the pair met via Zoom to compare drawing codes and discuss where there was a disparity to produce an agreed coding. Disparities were infrequent and mainly due to a missing code where more than one code applied to a diagram. Both coding pairs also met via Zoom, discussed their coding, and reached a consensus in the presence of other project members. This ensured that the overall coding process was consistent and reliable.

Phase 4: charting. The charting stage involves summarising the indexed and sorted data into a coherent analytical picture, which is communicated in the results.

573 **Phase 5: mapping and interpretation.** The final stage in the process moves towards pulling key aspects of
 574 data across the data set in order to understand and interpret it as a whole. Further description, clarifying concepts,
 575 representing the range and depth of data, establishing relationships, and developing explanations are all practices
 576 relevant for this stage [71].
 577

578 We note that Phase 3 was iterated a few times to clarify the themes and reach a consensus within and between
 579 the researchers. We provide a more comprehensive description of the four themes that emerged from this process in
 580 Table 12 in the Appendix. Section 4.1 reports on the outcome of the drawing analysis.
 581

582 Table 3. Drawing Coding Categories
 583

584 Category 1	Mirroring the depicted scenario directly in the drawing
585 Category 2	Imagining potential next steps and bad actor actions, more than the presented scenario suggests
586 Category 3	Identifying potential account compromise (loss of credentials)
587 Category 4	Identifying sensitive & personal identifying information leakage

589 3.4.2 Transcript Analysis Method.

590 Transcript analysis included all workshops and used audio recordings that contain two main types of interaction: (1)
 591 children who volunteered to explain their drawings, and (2) researcher-child unstructured interviews at the end of
 592 drawing sessions.
 593

594 Not all children participated in unstructured interviews due to time constraints. We could not observe the selection
 595 process, as it was teacher-led and cameras were switched off. During the unstructured interviews, children were
 596 questioned about their understanding of the scenarios, the perceived consequences of clicking, their familiarity with the
 597 scenario context and the sources of their knowledge. Discussions were typically guided by the researcher's interpretation
 598 of what would be meaningful to the child and proceeded with further questions to explore the child's responses and
 599 comments.
 600

601 The transcript analysis used reflexive thematic analysis (RTA) [8, 9, 15] led by one researcher's interpretive analysis.
 602 The initial coding process by the lead researcher was then discussed amongst the research group. Four researchers in
 603 total (the lead plus three others) sense-checked ideas by exploring multiple assumptions, interpretations, and meanings
 604 of the data, following a collaborative and reflexive approach.
 605

606 The transcriptions were analysed with a paradigmatic framework of interpretivism and constructivism, reflecting on
 607 the children's own accounts of their attitudes, opinions, and experiences as faithfully as possible while also accounting
 608 for the reflexive influence of researcher interpretations. Our ability to identify what we saw in the data was informed by
 609 existing concepts, our own knowledge of the literature, and the drawing analysis framework. Hence, while the analysis
 610 was dominantly inductive, a degree of deductive analysis was employed to ensure that the open coding contributed to
 611 producing themes that were meaningful to the research questions.
 612

613 Both semantic and latent coding were utilised. No attempt was made to prioritise semantic coding over latent coding
 614 or vice-versa. Rather, semantic codes were produced when meaningful semantic information was interpreted, and latent
 615 codes were produced when meaningful latent information was interpreted. As such, any item of information could
 616 be double-coded in accordance with the semantic meaning communicated by the respondent and the latent meaning
 617 interpreted by the researcher.
 618

619 The analysis was carried out following the 6 phases of RTA:
 620

621 **Phase 1: familiarisation.** The main researcher participated in all the workshops and carried out or listened to all
 622
 623
 624

interviews. Some preliminary notes were taken after each workshop in collaboration with all the researchers at the given workshop.

Phase 2: generation of initial codes. The preliminary iteration of coding was conducted using the 'comments' function in Microsoft Word. This allowed codes to be noted in the side margin while also highlighting the area of text assigned to each respective code. Multiple comments were used for double coding. The initial codes were discussed with four other researchers.

Phase 3: generating themes. A Microsoft Excel spreadsheet was established to bring together all codes from all workshops for each scenario. The coded data was reviewed and analysed as to how different codes may be combined according to shared meanings to form themes.

Phase 4: review of themes. A thematic map is created from the review of themes.

Phase 5: definition and naming of themes. Four researchers reviewed themes based on the underlying data, following a collaborative and reflexive approach, and finalised the names and definitions.

Phase 6: reporting. Section 4.2 reports on the outcome of the RTA.

4 RESULTS

4.1 Drawing Analysis

In total, we analysed 468 drawings. Our analysis revealed that 144 (31%) mirrored our scenario back to us (Category 1 in Table 3), which indicates a non-intrusive reading of the scenario or the user giving access to information (e.g., an email address or biometrics) in a tit-for-tat exchange. 246 (53%) respondents began to imagine more than what was presented in any given scenario (Category 2). These responses identified that information was being tracked or collected via invasive means without user input (e.g., location services, information accessing). 164 (35%) participants began to identify what might be lost in terms of security or privacy in response to the scenario (e.g., theft of credentials or account details) (Category 3). 145 (31%) participants depicted worst-case scenarios where sensitive personally identifiable information (e.g., bank details, postcodes, date of birth, or phone number) was leaked (Category 4). In their drawings, 274 (59%) participants mentioned some form of the potential consequence of a given scenario. We can now consider drawings related to each scenario in turn.

The naming convention for drawings specifically mentioned in this Section is: 'WorkShop i' - 'Participant j' : WSi-Pj. See Figures 8 to 12 in the Appendix for drawings.

4.1.1 Scenario 1 – Privacy Zuckering.

Table 4 depicts all the coding of the categories in this scenario's drawings. Figure 2 provides one example of a drawing. Drawing WS3-P23 (Figure 8a) illustrates category 1 as it mirrors the scenario: one screen showing an eye and the text 'free gems', 'scanning...', 'camera opened'. Category 2 applies to drawing WS7-P18 (Figure 8c) where there is a strong reference to baiting a trap, including the pricing structure priming user action. The image shows one busy screen, a mid-game scenario with more credits needed 'oh great do you need gems', then three price options and free gems for retinal scan, the user clicking yes, and finally an image of an unhappy-looking person. Category 3 is applicable to drawing WS4-P10 (Figure 8d), which identifies account compromise as it shows a single mobile screen with the text 'lost connection your phone is being tracked'. Category 4 is reflected in drawing WS7-P08 (Figure 8e) where the child is aware of spyware, in a literal sense, as the image shows themselves on a screen with the text "they could see you through your camera".

Multiple categories were often coded for the drawings. Responses to Scenario 1 were coded more frequently in categories one and two, and less frequently in categories three and four, across the 133 drawings. In essence, many children either mirrored the scenario in their drawings or imagined deceptive attempts.

In some drawings, there was a sense of providing more information about themselves to get more gems, e.g. in drawing WS6-P12 (Figure 8b), the first image depicted the eye and eyebrow with narrating text “you open your camera and scan your eye for the gems”, the second image depicted lips with narrating text “You get the gems, and they offer you more gems for another body part”, and the final image depicted a head with smiling face and short hair and text written asking questions of user “D.O.B.: ??? Where you live: ??? Hobbies: ???”. This drawing was coded as illustrating categories 1, 2, and 4.

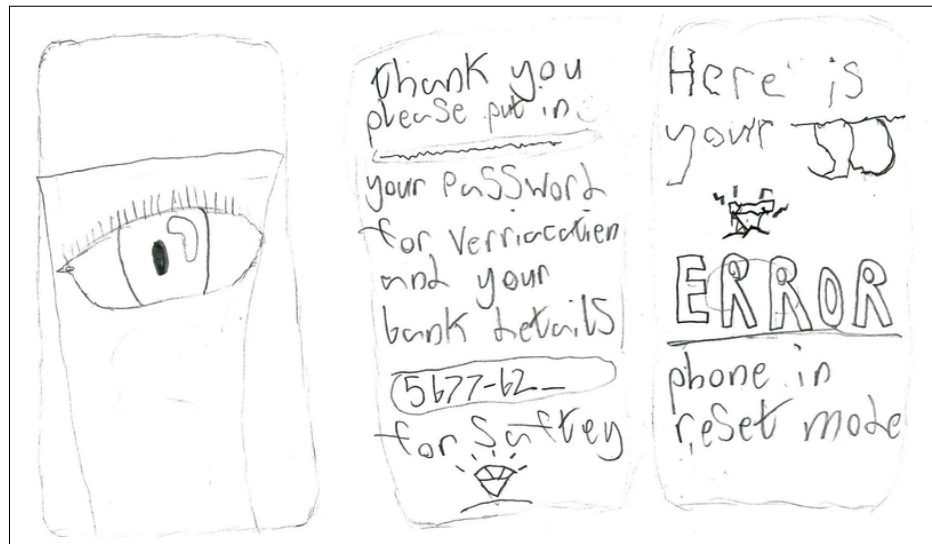


Fig. 2. Example Scenario 1 Drawing (WS5-P12)

Some did imagine undesirable consequences, coded in 85 of the 133 drawings. The nature of these consequences varied considerably, but ‘tracking’ was most commonly cited. Some thought the eye scan could lead to unintended and unwanted consequences, such as ID or card theft or even non-digital crimes (e.g., burglary - WS3-P08 - Figure 12c). Some expressed safety concerns related to someone being able to identify their school from the scan and coming to their homes with violent intentions (see WS3-P19 in Figure 12d). There were 6 examples where children anticipated that the camera would take face and body images rather than simply an eye scan (WS7-P08 - Figure 8e). Many of the children were suspicious but unclear about the nature of the threat. (See Table 8 for a full range of mentioned consequences.)

Moreover, 22/133 (17%) drawings construed unrealistic consequences, considering this scenario to depict a ‘Bait & Switch’ attack (WS6-P12 - Figure 8b). The children did not depict any indicators that could be linked to loss of privacy, suggesting they may not have understood the value of their biometric or the risks of giving away their eye scan.

4.1.2 Scenario 2 – Bait & Switch.

Table 5 depicts all the coding of the categories in this scenario’s drawings. Figure 3 provides an example of a drawing for this scenario. Responses to Scenario 2 were coded most often as in Category 2 (54% – 66 of 122). It is clear that the

Table 4. Coding of Scenario 1 responses (n=codes)

n	Workshop	Mirroring Category 1	Next Steps Category 2	Potential Compromise Category 3	Info Leakage Category 4
11	WS2	6	2	2	1
53	WS3	14	16	4	19
35	WS4	8	14	11	2
37	WS5	8	12	4	13
22	WS6	9	6	3	4
30	WS7	11	9	2	8
188	Total	56	59	26	47

participants could imagine more than what the provided scenario showed. A typical response was noted in drawing WS5-P17 (Figure 9a), where the child imagined being forced to download a game before being offered the free Robux. It is clear that the majority spotted this dark pattern.

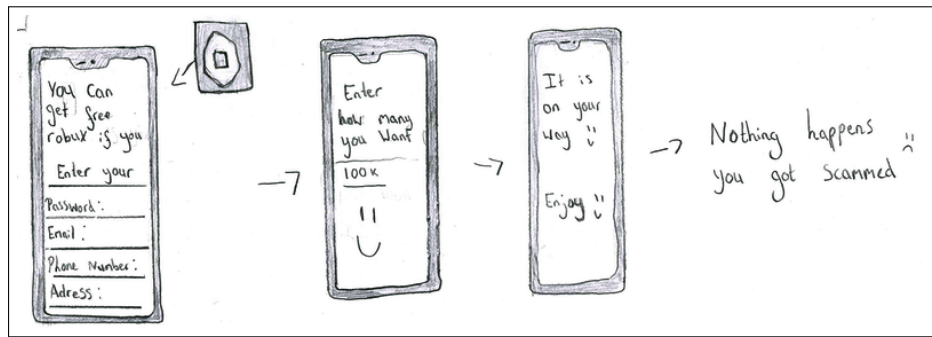


Fig. 3. Example Scenario 2 Drawing (WS2-P06)

Imagined consequences included the compromise of accounts (Category 3) and the leaking of personal information (Category 4) (Table 9). Typical drawings included WS6-P08 (Figure 9c; Category 3), who drew a credentials form (username and password) as a first step after clicking on the scenario, and WS5-P11 (Figure 9b), who mentioned their bank details being requested. In general, we note that bank/financial details data collection was a common theme referring to data leakage across Scenario 2.

Category 1, that is, mirroring the scenario back without imagining deceptive practices (see WS7-P19 (Figure 12g), WS3-P23 (Figure 8a), had the lowest occurrence in Scenario 2, with 16% (20 of 122) participants. We also note that Category 1, when present in this scenario, was followed by Category 3 or 4 or a consequence, such as WS3-P05 (Figure 9f), who depicted: (1) clicking on free Robux, (2) waiting time for Robux arrival, and (3) phone gets hacked.

Consequences across Scenario 2 were named in 76 of the 122 responses (accounting for 62%). Being hacked as a result of the scenario was named most often, across 34 responses, such as WS3-P21 (Figure 9d) "you have been hacked" and a smiling face, while money theft and/or emptied bank account was in the second position, with 21 instances, such as WS5-P22 (Figure 9e) "You have been no money scammed". We note that hacking often referred to bank accounts or Robux accounts with the additional consequence of money theft or Robux (account) theft. (See Table 8)

Workshop 3 scored highest in the consequences of hacking and money theft, followed by Workshop 5. These two workshops had the highest number of participants.

Table 5. Coding of Scenario 2 responses (n=codes)

n	Workshop	Mirroring Category 1	Next Steps Category 2	Potential Compromise Category 3	Info Leakage Category 4
11	WS2	6	2	2	1
52	WS3	8	20	10	14
30	WS4	1	16	5	8
40	WS5	2	7	15	16
17	WS6	3	9	2	3
34	WS7	1	18	13	2
184	Total	21	72	47	44

4.1.3 Scenario 3 – Confirm Shaming.

Table 6 depicts all the coding of the categories in this scenario’s drawings. Figure 4 provides an example of a drawing for this scenario. Responses to Scenario 3 were coded most frequently in categories one and two and less frequently in categories three and four. 60/118 (51%) participants recognised or imagined deception-based data access (Category 2), and 47/118 mirrored the scenario back (Category 1). 35/118 (30%) illustrated the consequences of a leaked account (Category 3), and 22/118 (19%) participants expressed Leaked PII (Category 4).

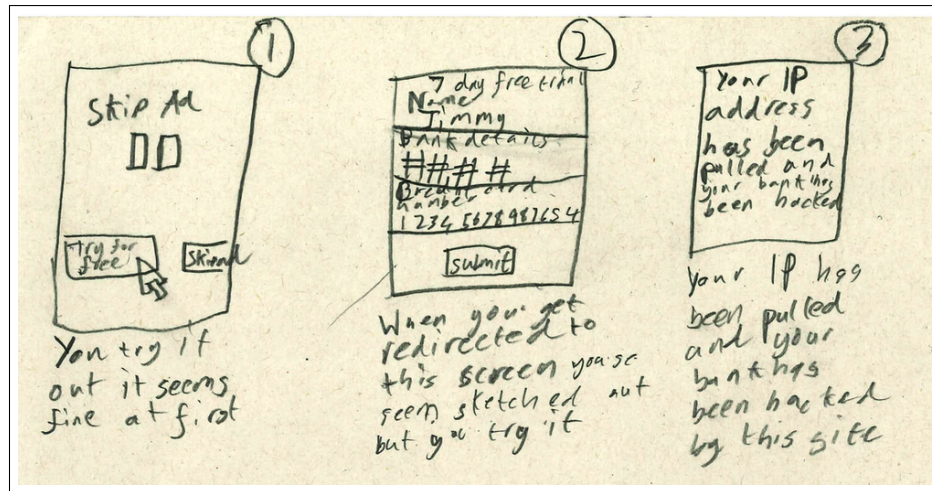


Fig. 4. Example Scenario 3 Drawing (WS6-P3)

Consequences imagined from the scenario were named in 60 of the 118 samples (See Table 8). The most commonly named consequence was ‘hacking’ (20), followed by ‘financial loss’ (11), ‘virus’ (9) and ‘scam’ (8).

The quantitative analysis showed some variability across workshops, where WS3 shows a disproportionately high number of Category 1 codes (i.e. mirroring back scenario) and fewer Category 2 codes. Other workshops tended to indicate more Category 2 codes than Category 1 codes (WS3-P04 - Figure 10a). WS5 appears to have had a much larger number of Category 2 coded drawings (e.g., WS5-P05 - Figure 10b). The trend, overall, seems to be that around half of the children indicated Category 2 codes (where they start to imagine more than what is presented). There were 23 examples in which anticipated requests for further details led to unwanted consequences. In terms of consequences, a

quarter of participants identified what is being lost in terms of security/privacy (Category 3), and a quarter identified hacking of other sensitive/personally identifiable data (Category 4) - see WS3-P12 (Figure 10c). These indicate that some of the children were cautious, wary, and suspicious of unknown situations.

Table 6. Coding of Scenario 3 responses (n=codes)

n	Workshop	Mirroring Category 1	Next Steps Category 2	Potential Compromise Category 3	Info Leakage Category 4
9	WS2	1	5	2	1
48	WS3	21	8	11	8
32	WS4	9	12	11	0
31	WS5	2	18	2	9
17	WS6	3	9	2	3
27	WS7	11	8	7	1
164	Total	47	60	35	22

4.1.4 Scenario 4 – Browser Warning.

Table 7 depicts all the coding of the categories in this scenario’s drawings. Figure 5 provides an example of a drawing

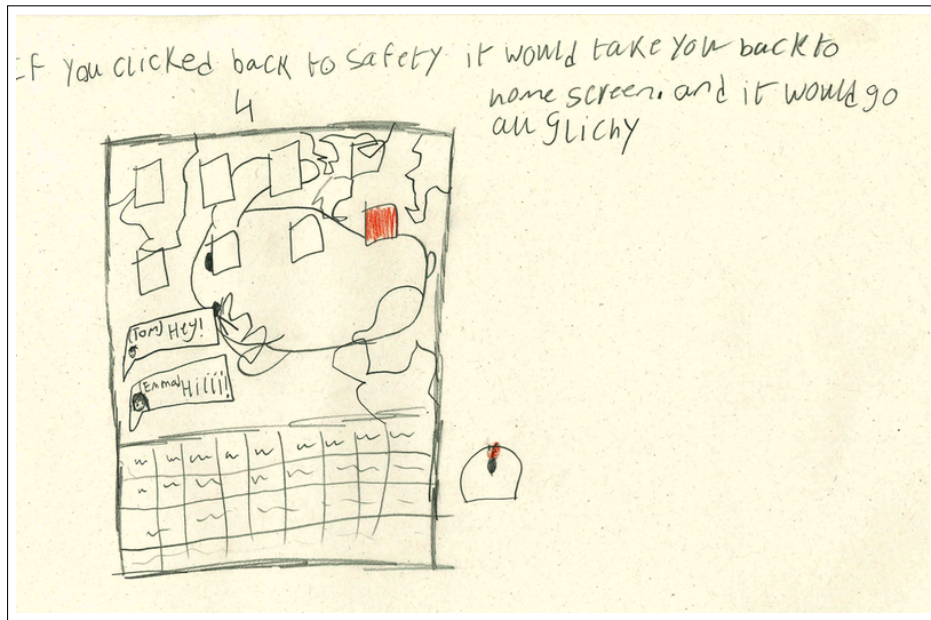


Fig. 5. Example Scenario 4 Drawing (WS4-P15)

for this scenario, where the child imagines that even if the ‘Back to Safety’ option is chosen, the computer would still be “glitchy”. Responses to Scenario 4 were coded more frequently in categories two and three (WS7-P16 - Figure 8c) and less frequently in categories one and four. Participants recognised or imagined deception-based data access (Category 2) and leaked account consequences (Category 3) more commonly, at 44/98 (45%) and 45/98 (46%), respectively. Mirroring

the scenario back was less prevalent for this scenario, with only 27/98 (28%) respondents doing so (WS7-P09 - Figure 11a).

Table 7. Coding of Scenario 4 responses (n=codes)

n	Workshop	Mirroring Category 1	Next Steps Category 2	Potential Compromise Category 3	Info Leakage Category 4
6	WS2	0	4	0	2
45	WS3	13	10	11	11
23	WS4	4	7	10	2
20	WS5	2	13	3	2
18	WS6	1	3	8	6
28	WS7	7	7	13	1
140	Total	27	44	45	24

Imagined consequences included 'hacked' (29), 'virus', or 'bug' (11). Bear in mind that this scenario is not a dark pattern and actually tries to warn the user about a potentially fake or harmful website being accessed if they continue.

4.1.5 Cross-Scenario Comparison.

Table 8 lists the consequences mentioned by the children across all scenarios. They demonstrate a familiarity with the terms 'hacked' and 'scam'. In terms of specific losses, they often mentioned credit card theft. Privacy-related consequences were rarely mentioned, and while safety-related consequences were mentioned, these were not widely cited by the children.

Scenario 1 was clearly evocative, with many potential consequences being mentioned, mostly related to cybersecurity. This scenario elicited many imagined consequences, most of which were unrealistic (e.g., an eye scan being used to find the child's home address (WS3-P08 Figure 12c). Amongst all scenarios, Scenario 2 (Bait & Switch) had the highest percentage of participants: (i) imagining next steps beyond what is provided in the scenario (Category 2); (ii) mentioning account compromise, which included both compromise of Robux and bank accounts (Category 3); and (iii) pointing to personal data loss (Category 4)(WS3-P30 - Figure 12f).

While Scenario 1 depicted Privacy Zuckering, that is, being tricked into disclosing more personal information than is wise, we found that Scenario 1 had the highest % of drawings overall that merely mirrored the scenario in the drawing (Category 1, 56/133). Furthermore, in Scenario 1, this was followed by participants starting to imagine giving access to more of their own information via the eye scan (Category 2, 60/133). There were fewer depictions of personal data being released and of identification and sensitive types of data being compromised (Category 4) and account information (Category 3).

Children miscategorised most scenarios as 'Bait & Switch', even when only one was this kind of dark pattern in reality. Hence, while they were wary, they were not able to distinguish one dark pattern from another nor to identify the one benign scenario in the mix (Scenario 4).

Table 9 demonstrates stark differences between the children participating in the different workshops. WS2 participants only came up with 10 consequences, while WS3's participants came up with 87. Many factors might be influential here with parental knowledge [62], levels of deprivation [61] and children not yet having developed the required skills [40] being but a few. Still, it is interesting to see such differences even at such a young age, all of which demonstrate the need to ensure that children do indeed receive online deception-related education as they start to operate autonomously online.

Table 8. Frequency analysis of named consequences per scenario (n=mentions)

n	Consequence	Number per Scenario			
		Mirroring Scenario 1	Next Steps Scenario 2	Potential Compromise Scenario 3	Info Leakage Scenario 4
NON-SPECIFIC					
100	Hacked	17	34	20	29
33	Scam	8	12	8	5
1	Spam	1			
1	Warning message				1
SPECIFIC LOSS (INFO/FINANCIAL)					
4	Info deleted	3			1
21	ID theft	14	1	3	3
51	Account/credit card theft	12	21	11	7
16	Account compromised		10 (Robux account)		6
DEVICE COMPROMISED					
34	Virus/malware	9	9	11	5
12	Device shut down/blank screen			6	6
2	Glitch/error				2
PRIVACY LOSS					
12	Track location	9	1		2
7	Taking extra camera images	6			1
2	Friends contacted				2
3	Info shared online	2			1
PERSONAL SAFETY COMPROMISED					
3	Burglary	3			
5	Physical attack	1			4
OTHER					
2	Illegible			1	1
TOTAL					
309		85	88	60	76

4.2 Transcript Analysis

The transcripts from the workshops represented 22-29 different voices for each scenario, which corresponds to roughly 20% of all drawings. In the transcripts, each speaker was given a different participant label. The participant labelling convention for transcripts specifically mentioned in this Section is 'WorkShop i' 'Scenario j' 'Child k' : WSiSjCk. Table 10 lists the labels for each interviewed participant and presents the total interviews at each workshop, as well as the total. However, as we could not observe the classroom while the children spoke, there is a remote chance that some children may have spoken multiple times (the facilitating teacher ensured turn taking, and we were not able to see children). Hence, a single label may not always refer to a distinct individual.

Even though we could not conduct an interview for every drawing, the transcripts provided a useful sample of what children thought of how and why a benign or malicious action would occur in response to a click. The transcripts were

Table 9. Total drawings/responses naming at least one consequence, by workshop scenario

n	n Workshop	# Drawings naming at least one consequence			
		Mirroring Scenario 1	Next Steps Scenario 2	Potential Compromise Scenario 3	Info Leakage Scenario 4
10	WS2	2	3	3	2
87	WS3	24	21	19	23
50	WS4	13	13	11	13
78	WS5	28	18	15	17
25	WS6	6	5	5	9
46	WS7	12	16	7	11
296	Total	85	76	60	75

Table 10. Workshops and Participant Labels for Each Child Speaking for a Particular Scenario ('WorkShop i' 'Scenario j' 'Child k')

Workshop	Interviewed Child Participant Labels				Total Labels per Workshop
WS1	WS1S1C1-C3	WS1S2C1-C5	WS1S3C1-C5	WS1S4C1	14
WS2	WS2S1C1-C3	WS2S2C1-C6	WS2S3C1-C3	WS2S4C1-C5	17
WS3	WS3S1C1-C3	WS3S2C1-C6	WS3S3C1-C3	WS3S4C1-C2	14
WS4	WS4S1C1-C4	WS4S2C1-C5	WS4S3C1-C4	WS4S4C1-C3	16
WS5	WS5S1C1-C2	WS5S2C1-C2	WS5S3C1-C5	WS5S4C1-C2	11
WS6	WS6S1C1-C5	WS6S2C1-C3	WS6S3C1-C3	WS6S4C1-C3	14
WS7	WS7S1C1-C2	WS7S2C1-C2	WS7S3C1-C4	WS7S4C1-C4	12
Total:					98

also richer in terms of children's reactions to the scenarios and their expectations of online safety, security and support. Therefore, the sample was useful in revealing the varying views and understandings of privacy and security.

Three themes were produced by organising codes around a relative core commonality interpreted from the data: (1) dark pattern perceptions; (2) online behaviours; and (3) security knowledge, which comprised nine sub-themes (see Figure 6). Table 11 in the Appendix presents themes, sub-themes, codes, and additional examples from transcripts. The first theme characterises the ways in which participants described their drawings in relation to dark patterns. Dark pattern perceptions, unsurprisingly the most substantive points of discussion, illustrate the ways in which participants were found to understand the four scenarios in terms of *nefarious actors*, *actions*, and *consequences*. Moreover, participants' discussions on online behaviours were analysed and fell into *cautious* or *risky practices*. Beyond the scenarios themselves, and the initial scope of the project, participants also demonstrated some security knowledge and spoke about the sources of their knowledge. Collectively, these themes paint a picture of the way children in the selected age group at the participating schools viewed the online environment and guide themselves through it.

Theme 1: Dark pattern perceptions:

In discussing their drawings of dark patterns embedded within the four scenarios, children included four concepts: (1) leaked data via user interactions or snooping, (2) nefarious actors, (3) actions, and (4) consequences. The children gave examples of *different data* they may be giving away either directly, e.g., filling out a form, or indirectly, e.g., through being snooped on. *Actors* relate to participants themselves as well as others within a given scenario. *Actions* pertain to what happens within, and in response to, a scenario. Expected *consequences*, the majority of the time, referred to negative outcomes, although, in some instances, children indicated that they thought nothing would happen.

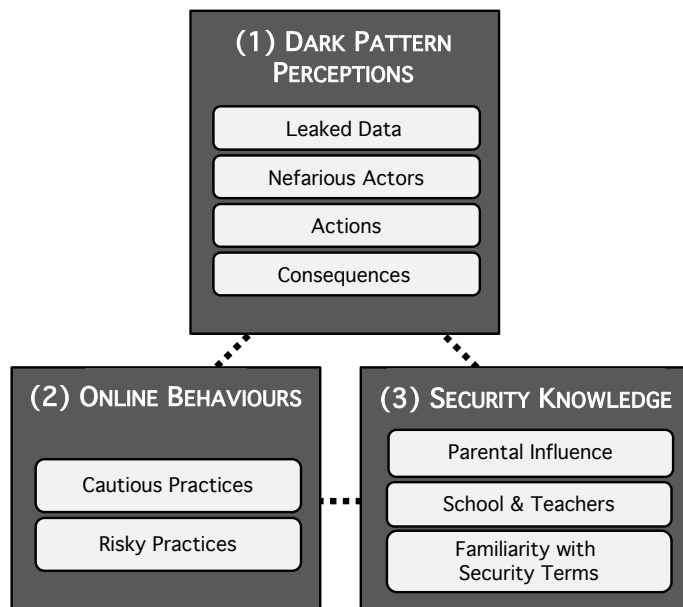


Fig. 6. Thematic map demonstrating 3 themes and 9 sub-themes

Leaked data via user-interaction or snooping. This sub-theme includes discussions on what data is shared by the user by, e.g. filling in a form versus what data may be leaked even if the child does not enter any information. While children were vague, at times, in terms of what data would be at risk, resorting to ‘more information’, ‘your details’, and ‘private data’, they frequently pointed to being asked for email, password, and bank details. In addition, name, address/postcode, phone number, gender, birth date, photos, and important files were mentioned. However, it was clear that the children did not always know why this information was being requested: “*not sure why they want things like your phone numbers, just know that they do*” (WS4S1C4). For Scenarios 2 and 4, some children mentioned friends or friend lists, which we assume to be due to social media use triggered by the scenario, e.g., “*Then they get access to your friends list and they tell them you’ve gone to the website*” (WS5S4C1). This demonstrated an awareness of the variety of data that may be at risk (‘personal data’ mentioned by European Union’s General Data Protection Regulation). In addition, they understood that others were likely to be impacted (friends being hacked as well) as a result of their falling victim to a dark pattern enabled exploit.

Nefarious Actors. This sub-theme includes descriptions of who is involved in the scenario, also the honesty levels and capabilities of hackers. Children spoke about the scenarios through the lens of their own experiences or without pointing to a particular actor. They used statements like ‘I/you allow’, ‘I/you end up’, ‘I/you need to’, or ‘I/you have to’ when describing their drawings of scenarios.

However, when referring to the human behind the scenario or dark patterns, ‘hackers’ was the overwhelming term used. A gendered association was made with hackers, who were often referenced as ‘a guy’ or ‘he’; very rarely ‘she’ was used. In one instance, the offending party was named “*people who want Gucci*” (WS3S3C1). In contrast to human actors, software was referenced as a non-human actor ‘it’. These actors ‘tell’, ‘want’, ‘direct’, ‘pressure’, ‘urge’, ‘guilt-trip’, or ‘bribe’, indicating different levels of manipulation by urging input or action.

1093 Children also imagined different actor capabilities and sophistication. Hackers were understood to be able to use ‘high
1094 tech’ strategies motivated by financial gain. For example, one child said, “*the hackers have strong technology*” (WS3S2C6).
1095 Motivations were often linked to financial gain. One participant aptly captured the capabilities of hackers when they
1096 remarked, “*some people saying they’re not putting the details in money is getting taken out anyway*” (WS5S3C4).
1097

1098 **Actions.** This sub-theme captures the children’s descriptions of different steps involved in the scenario leading to
1099 deception. For example, in the Privacy Zuckering scenario, the children imagined different ways a device might ‘scan
1100 your eye’. Some imagined their torso being scanned, some their face or just their eye, thus demonstrating different levels
1101 of privacy invasion. For example, one participant shared, “*They say they want your eye, but then they are going to take*
1102 *your whole face [...] I’m going to have to zoom in so much, and then they can see my whole face*” (WS6S1C2). One child
1103 commented “*He says it’s bribing you to take a picture of yourself, and that’s all. You won’t get the gems it has promised.*
1104 *The game goes as normal*” (WS1S1C2) without mentioning any loss of privacy due to their eye scan being taken. Some
1105 children considered the scenario more a ‘Bait & Switch’ - “*Direct you to take a picture of your eye, and then that will*
1106 *basically hack your phone*” (WS1S1C1). or “*If you click this button a virus gets onto the computer. It will ask for your bank*
1107 *details*” (WS5S1C2). In the ‘Bait & Switch’ scenario, children mostly expected to be taken to a fake website and asked
1108 to enter a username/email and password, explaining: “*So I clicked it and it took me to this website and it says please enter*
1109 *your Roblox username*” (WS6S2C1), or “*And they click on a website called, notascam dot yay*” (WS7S2C2), which signals
1110 the deceit with the name choice in the example given for the website. Some children only loosely identified the scenario
1111 without mentioning visiting a fake website but expecting user information to be leaked or Roblox to be stolen (e.g., “*So*
1112 *once you click on it, it will ask you for all your information*” (WS1S2C5) or “*It’s a scamming tool. It will steal your Roblox*”
1113 (WS4S2C4).
1114
1115
1116
1117

1118 In the third ‘Confirm Shaming’ scenario, many communicated they would skip the ad and nothing will happen: “*Just*
1119 *skip that and get your video on YouTube. That always comes up*” (WS6S3C1), “*So what happens is when you click this*
1120 *button, you continue to watch YouTube, and that’s pretty much it. And then you’ll be happy*” (WS1S3C1), “*nothing would*
1121 *really happen because I skipped a lot of ads before and nothing ever happened*” (WS2S3C2). One child said, “*It’s trying to*
1122 *guilt trip you into watching the ad*” (WS7S3C1), describing the dark pattern aptly. However, some also classified this as a
1123 ‘Bait & Switch’: “*I think when you press skip, that’ll take you to another tab of YouTube, but it’s fake YouTube*” (WS2S3C1).
1124
1125

1126 In the fourth scenario, some children recognised the legitimate warning: “*Just click to safety and not get a malware*
1127 *virus*” (WS3S4C2). Typically, the children assumed ‘Advance Anyway’ would result in a ‘Bait & Switch’ style redirection:
1128 “*So you click on advance anyway, and this isn’t the actual Instagram. It’s a knock-off Instagram*” (WS6S4C3).
1129

1130 These descriptions of different actions showed how children assessed a deceptive scenario and the risks attached to
1131 it, giving us more insights into RQ1 and RQ2. The children, the majority of the time, assumed malicious intent, even
1132 with the legitimate warning, and anticipated a ‘Bait & Switch’ style exploit behind them.

1133 **Consequences.** The children mentioned several consequences, which we roughly categorised into five areas: (1)
1134 offline harm, (2) leaked data shared with others or used online to scam or spam the victim, their friends, or others, (3)
1135 bank account hacked and financial loss, (4) web account(s) hacked, (5) device hacked or infected with a virus.
1136

1137 An example of offline harm (1) was communicated in response to the Privacy Zuckering scenario, where one child
1138 connected their online behaviour with physical safety and risk to imagine, “*The camera can see your school uniform.*
1139 *Then they will see my badge, and then they can come to school*” (WS6S1C2). Others shared, “*They could track you down*
1140 *and like maybe steal stuff in your house*” (WS3S1C3), or “[*Data leak would be horrible*] ‘*cause you don’t really want*
1141 *someone to come to your address*” (WS7S4C1).
1142
1143
1144

1145 One example of the shared or leaked data online to scam the victim (2) was also shared in response to the Privacy
1146 Zuckering scenario: one child said, "*They could maybe use your face and then just start spreading fake rumours*" (WS7S1C2).
1147 Another commented, "*When you click that, it will take you to one of their applications and those guys spam you*" (WS4S3C2).
1148

1149 While the Privacy Zuckering scenario yielded more offline harm responses, the other three scenarios overwhelmingly
1150 resulted in either a financial loss or account or device hacking. Financial losses (3) were imagined to occur through
1151 a bank account being hacked, e.g., "*then you get an email from your bank account. There's been an error, and all your*
1152 *money is being spent*" (WS3S2C5) or "*they can hack into your phone and log into your robux account and use your email to*
1153 *buy a lot of things*" (WS1S2C3).
1154

1155 Accounts were hacked (4) either by entering credentials and having them stolen or simply by clicking the wrong
1156 button. Another child discussed this in terms of being conned into entering login credentials to a fake website: "*Then*
1157 *you tried to log back in [YouTube], they would have already changed your password*" (WS2S4C1).
1158

1159 Device hacking (5) was understood to occur in different ways, from taking control of a camera or software or shutting
1160 a phone down. In response to Scenario 4, one child (WS3S4C1) imagined that the example URL, 'Instagram. con' would
1161 enable a hacker to see their photos, continue using their camera, or watch their victims. This child explained: "*If you*
1162 *click on advance anyway, hackers will be able to take over your PC and move your mouse to their advantage, which is not*
1163 *recommended. Then they could be watching you through your phone camera on your phone*". Most interestingly, a few
1164 suspected this scenario to be a 'double bluff' (WS6S4C1): "*The back to safety button would also give you a virus. Either*
1165 *way, you are going to get a virus*" (WS4S4C3).
1166
1167
1168

1169 Theme 2: Online behaviours

1170 The behaviours in response to the four scenarios were discussed in two primary ways: cautious and risky behaviours.

1171 **Cautious behaviours** The children exhibited cautious behaviours, such as not engaging, asking a parent, or
1172 reporting suspicious activity. Several examples were communicated (e.g., leaving a game, not inputting any details,
1173 closing/refreshing an app, or deleting it). For example, "*I wouldn't put my details in*" (WS5S2C2) or "*I would just close the*
1174 *app and delete it*" (WS7S4C2).
1175

1176 Another common cautious behaviour was to ask a parent, exclusively a mother. Indeed, fathers were not mentioned
1177 during the study. One participant said, "*I always ask my mum if I can get games [and that] my mum has to accept*"
1178 (WS6S1C5). Mums also put rules in place for online behaviour: "*If it's social [media], I have to ask her*" (WS6S1C5); an
1179 alternative was to use family sharing of app and content producers (e.g., "*My mum and I have a thing called family*
1180 *sharing on my devices at home. On the App Store, I have to type in my password and then it says ask and then it sends a*
1181 *notification to my mum's phone and she has to then accept it and type her password in*" (WS6S2C3). The children would
1182 report suspicious activity in certain instances. In response to Scenario 2, one child would report the behaviour to a
1183 reputable website, "*report like on YouTube*" (WS6S2C3); and another shared, "*the best way to prevent this by reporting in*
1184 *the person or contacting Robux. So he would not be able to scam anyone else*" (WS3S2C4).
1185
1186
1187

1188 **Risky behaviours.** Several risky behaviours were mentioned, including not reading the 'Terms of Service', ignoring
1189 age-related warnings, or brand bias. Similar to adults, these children did not read the terms of service: "*I can't be*
1190 *bothered. Click the accept button*" (WS6S1C4), another providing a rationale: "*Because it's too long. No one wants to read*
1191 *all of it*" (WS2S1C3).
1192

1193 Age-related content warnings were raised as items to ignore, as though barriers to playing a game. One participant
1194 said, for example, "*It's just car racing*" (WS6S4C1) and thus not an important/relevant warning. Finally, some participants
1195 exhibited a degree of brand trust that influenced or underpinned risky behaviour. In response to Scenario 2, one child
1196

1197 said, “*You might think: it’s YouTube. It’s a big company. There won’t be any problems*” (WS6S3C1).
1198

1199 **Theme 3: Security Knowledge:**
1200

1201 An important extension of the scenario-based discussion was where children had learned about navigating online risks,
1202 the challenges, and dangers. Whilst we did not set out to elicit this information, it became a relevant part of discussions
1203 and provided insights into knowledge gathering and patterns of online behaviour. Children demonstrated familiarity
1204 with terminology and shared two primary sources of information: (1) schools & teachers, and (2) parents.
1205

1206 **Parental influence.** Parents were primary sources of knowledge. Some children asked their parents when encoun-
1207 tering certain scenarios online. One child said: “*I have asked my mum if it was like a scam or something; I just ask mum*
1208 *for advice about it*” (WS7S4C3), and another shared a personal experience at home with negative effects: “*My mum was*
1209 *hacked once [...] my mum always told me. Like, don’t trust things like that*” (WS7S2C1).
1210

1211 **Schools and teachers.** Generally, the children cited their school as the source of their security knowledge. Although
1212 they could not give details on definitive programs or training, they did say that they had been taught about online
1213 behaviours. This was aptly phrased by one child: “*I can’t remember what primary year it was, but we learned about it in*
1214 *school once as well*” (WS6S1C5). Teachers communicated the need for caution while online, which a child explained:
1215 “*And my teachers always said that they [don’t trust things like that] as well*” (WS7S2C1).
1216

1217 Interestingly, where the children used tablets (often iPads) at school, devices were considered safe. It was assumed that
1218 they could not be breached, as one child explained, “*[can they get into it?] not on our school iPads*” (WS6S1C4). Schools
1219 clearly informed children’s knowledge of internet safety, but there were noticeable gaps. The children and their teachers
1220 had less familiarity with the ‘Privacy Zuckering’ scenario. Hence, a few children questioned the motivation behind the
1221 request with queries such as “*why would they want to see you? no app has ever asked to scan my eye*” (WS6S1C1), “*cause*
1222 *they want the little kid’s face to do something [else]*” (WS6S1C3).
1223

1224 **Familiarity with terminology.** The children had awareness of dark patterns, using phrases, such as ‘dodgy’,
1225 ‘sketchy’, ‘fake’, ‘scam’, ‘spam’, ‘hack’, ‘virus’, ‘malware’, ‘trojan’, ‘remote PC’, or ‘dark web’, as well as safety icons,
1226 like the ‘browser lock icon’, which one person attributed to safety, as in, “*they’re safe websites with the lock at the side*”
1227 (WS4S2C5). The children did not mention more recent types of attacks involving, for example, ransomware or deep
1228 fakes.
1229
1230

1231 **5 DISCUSSION**
1232

1233 With respect to our chosen methodology, we designed the study to gather and analyse drawings as well as transcripts.
1234 However, it became clear, as we analysed both of these, that the combination of these gave us much richer insights than
1235 drawings on their own [60]. As such, even if we carry out further workshops with children in person, we will continue
1236 to collect both drawings and transcripts to explore different dimensions of the children’s mental models. By bringing
1237 together the findings from the transcript and drawing analysis, we are now able to answer the research questions posed
1238 in Section 3.
1239
1240

1241 **RQ1a** considers whether the children are able to spot dark patterns. There is strong evidence that they are indeed aware
1242 of the presence of bad actors and also of the fact that they are ‘up to no good’. However, their response to scenario 3
1243 demonstrates a tendency to classify all ‘sketchy’ scenarios as the ‘Bait & Switch’ dark pattern. This might be because
1244 they have already encountered this pattern online or because they have heard adults talking about the financial losses
1245 they have suffered due to being deceived by ‘Bait and Switch’ dark patterns.
1246
1247
1248

1249 This misclassification has two potential consequences. **Firstly**, they may miss other kinds of deceptive attempts.
1250 For example, those children who focused on 'Bait & Switch' in Scenario 1 would be less likely to realise that their
1251 eye biometrics would be captured and possibly sold to others. The **second** consequence is that they may construe
1252 unrealistic worst-case scenarios (e.g., WS4-P20 (Figure 12e) and WS3-P30 (Figure 12f)). For Scenario 1, some children
1253 envisaged an actor being able to identify their school from the webcam capture and come to their homes (WS3-P19 -
1254 Figure 12d). Similarly, in response to Scenario 3, which pushed users towards viewing an advert with a view to selling
1255 products or services, they imagined viruses and their devices being controlled by hackers (WS3-P05 - Figure 9f).
1256

1257 These findings align with the findings of Oates *et al.* [55] who found that children of this age were only starting to
1258 think about privacy in digital terms: making the transition from exclusively thinking about physical privacy at around
1259 age 10. Even so, similar to many adults, they do not seem to realise that their biometrics ought also to be kept private.
1260 Thinking that every dark pattern can potentially harm their devices and steal their information suggests that undue
1261 levels of anxiety experience while online, and being in a continuous state of suspicion is unhealthy. This is particularly
1262 true where the anxiety extends to concerns about their personal safety.
1263

1264 Our findings also confirm the arguments of Barnard-Wills in referring to e-safety education [4, p. 245]: "*Safety is*
1265 *a much more prominent concept than privacy, and privacy is never articulated as a stand-alone value, but only as an*
1266 *instrumental methodology or tactic for ensuring broader personal safety*" (WS3-P08 - Figure 12c). Moreover, as Marwick
1267 and boyd [49] point out, traditional individualistic conceptions of privacy might no longer be appropriate, especially in
1268 the social networking era, necessitating more innovation in this space. It is certainly evident that privacy education
1269 cannot be neglected [86].
1270

1271 **RQ1b** queries whether children can distinguish a genuine from a dark pattern scenario. We observed a somewhat exces-
1272 sive wariness that led them to suspect everything, even a genuine warning (WS4-P06 - Figure 12h). This is evidenced by
1273 the surprisingly high number of children drawing Scenario 4 who thought that clicking on the 'Back-to-Safety' button
1274 would also lead to them being hacked or getting a virus (e.g., WS2-P01 - Figure 12j). As such, the answer to **RQ1b** is in
1275 the negative.
1276

1277 **RQ2**, considers how well 11-12 year old children understand the motivations of bad actors using dark patterns.
1278 The findings suggest that most of the children had superficial and speculative models of bad actors' motivations. The
1279 references to financial theft mirror the 'burglar' model identified in the studies of adult users by Wash [91]. The high
1280 incidence of references to gratuitous hacking without a clear gainful motive suggests a similar model to the 'mischief'
1281 and 'vandal' models from that study. This is reinforced by several images and quotes in the drawings referencing a
1282 gloating or taunting actor.
1283

1284 As such, the answer to **RQ2** is that their understanding of motivations are sometimes incorrect (credential theft
1285 instead of privacy invasion) or exaggerated (hackers coming to their homes). This suggests that education design for
1286 online safety could usefully include descriptions of typical actors and their associated motivation to provide children
1287 with better insights into *what* bad actors are trying to achieve, and not just the *way* they try to achieve it.
1288

1289 **RQ3** considers how well 11-12 year old children understand the actions of bad actors and the potential consequences.
1290 We found that the children frequently could, and did, identify a wide range of potential consequences of dark patterns.
1291 Some of these were anxiety-producing (personal safety issues: WS3-P29 - Figure 12i), or non-specific (scammed/hacked).
1292

1293

1294

1295

1296

1297

1298

1299

1300

1301 Yet, as discussed earlier in this section, the tendency to classify all scenarios as ‘Bait & Switch’ dark patterns means that
1302 sometimes there is a mismatch between the actual risk and the consequences the children construed.

1303 In talking about consequences, the children in our study often used outdated language: most talking about viruses
1304 with very little mention of ransomware. The frequent references to bank accounts being compromised and their money
1305 stolen suggests that they may have been listening to adults talking about compromises. Indeed Rader and Wash [66]
1306 found that credit card and identity theft were most widely known by the general public. It is likely that the adults in our
1307 participants’ lives do not keep up with the latest cybersecurity threats, which is understandable given the complexity
1308 and dynamism of the domain.
1309

1310
1311 As such, the answer to **RQ3** is that children have a limited or outdated understanding of actions of bad actors and
1312 the consequences of falling for their deceptive attempts.

1313 It is important to note that our findings are relevant to an HCI audience as they demonstrate that design for children
1314 online has a significant ethical agenda. Safety of child users is compromised if they are vulnerable to nudging that is
1315 counter to their interests or involves deception.
1316
1317

1318 5.1 Future Work

1319 Some insights emerged that can inform future studies:

1320
1321 **Children need more nuanced insights:** One child said, of Scenario 4: “*It’s probably a real warning*” (WS7S4C2).
1322 When the researcher asked why, he said: “*It’s hard to explain*”. It seems that children have been made aware of the
1323 presence of dark patterns but not really given enough information to help them distinguish between dark patterns and
1324 genuine warnings. They assume the presence of the one pattern they know a lot about: the ‘Bait & Switch’ leading to
1325 fear-based responses. One possibility would be to co-design card decks or serious games with different stakeholder
1326 groups to help develop this ability.
1327

1328
1329 **More realistic expectations of consequences:** We observed a tendency for children of this age group to conflate
1330 cybersecurity risks with threats to personal safety. This seems to infuse cybersecurity with a measure of anxiety, which
1331 is bound to hamper their ability to distinguish good from bad. This generation of false positives suggests that while
1332 they are aware of threats, they do not have a sense of what cyber criminals might be trying to achieve, nor what their
1333 motivations might be [78]. We should be educating children to anticipate cyber criminal motivations and deceptions
1334 rather than focusing on specific attack and deception types.
1335

1336
1337 We need to do more research to find better ways to educate children rather than scaring them [67] and risking their
1338 construing personal safety consequences from cybersecurity threats.

1339
1340 **Understanding exploitation:** The children seemed to realise that permitting an eye scan could make them more
1341 likely to give more information: “*cause you think like oh they just asked me for my eye you might put in that information*
1342 *anyway*” (WS6S1C4). This child knows about the human need for consistency: having given one piece of information,
1343 this human tendency is likely to prompt further divulging of information. However, they did not seem to understand
1344 that their biometrics were valuable and that asking for these was another form of exploitation. They should be taught
1345 that they have the right to privacy and that their biometrics are personal data which should not easily be divulged.
1346

1347
1348 **Building resilience:** the children seemed to have a sense of doom and fatalism, feeling that if they were deceived,
1349 there would be no way of recovering. “*He has all your details, and you can’t do anything about that because you’ve*
1350 *agreed to all these terms of services*” (WS6S2C3). Hence, they ought to be given the skills to know how to recover from
1351 compromises and reassured that it is not impossible to recover and move on.
1352

1353 **The role of parents:** a number of children mentioned their parents (mums, actually) having given them advice
1354 about online deception possibilities. This suggests two directions. The first being that educating parents about dark
1355 patterns would be a good way of ensuring that the message is reaching children [62], which is especially beneficial
1356 since it would reach more children and also equip parents themselves to detect and resist dark patterns. The second is
1357 that it is essential to speak to both parents and their children, to gauge the influence of parents’ knowledge and efforts
1358 to educate their children in this respect.
1359

1360 5.2 Limitations

1361
1362 The contrived nature of our study, due to pandemic constraints, has led to limitations which we acknowledge. The
1363 primary limitation stemmed from the ethical protocols we elected to employ and accommodations we made resulting
1364 from COVID-19 (i.e., inability to collect data in person) and ethical sensitivities to online data collection with children.
1365 This limited data collection and analysis practices. Our preference for this work, which we suggest for future directions
1366 in this area, would have utilised in-person data collection wherein drawings could be matched with individuals and
1367 researchers could follow up consistently and comprehensively with the participants.
1368

1369 Due to the remote nature of the workshops and the decision not to collect video footage, we were unable to directly
1370 observe the classroom during the workshops, which had several implications: (1) some children might have spoken
1371 multiple times, with others remaining silent; (2) some children may have been able to see each other’s drawings while
1372 others may not have; (3) children may have discussed their drawings with one another while drawing, while others may
1373 not have. Similarly, while it would have added value to the data, we were not able to link the drawings to individual
1374 children. We considered this preferable to de-anonymising children, or asking teacher facilitators to issue children with
1375 anonymous codes, which could have been an error-prone process. Given our reliance on teachers as facilitators, we did
1376 not want to burden them with additional requirements that we would not be able to verify.
1377

1378 The children in Scottish schools mostly sit in groups of four at a single desk, which meant they could see each others’
1379 drawings as they drew. An element of groupthink might have influenced the outcomes of particular workshops. Factors
1380 such as perceived peer pressure [87] and FoMo (fear of missing out) may be particularly strong in teenage children [1],
1381 threatening to override trust considerations.
1382

1383 We presented the scenarios in the order we present them here, which might have biased perceptions. Finally, their
1384 teachers’ clarifications and instructions (as our facilitators) differed from workshop to workshop, and this, too, might
1385 have had an impact on the children’s drawings and responses. We also acknowledge that the instructions given to
1386 participants may have focused their attention on risk and the presence of bad actors more so than might be the case
1387 outside a workshop.
1388

1389 Finally, we do not claim that our findings generalise to other children of a similar age, even in Scotland. This was a
1390 snapshot of a few particular classes at a specific point in time. What our study *does* do is point the way forward for
1391 future work in this area, and highlights the need for more studies of this kind.
1392

1393 6 CONCLUSION & REFLECTION

1394 We set out to reveal 11-12 year-old children’s mental models of online dark patterns, and more specifically of the
1395 motivations of the bad actors who deploy these, and what the consequences would be if they were deceived, i.e. *data lost*,
1396 the *actors*, their *actions*, and the *consequences*. We carried out this study during pandemic lockdowns, which precluded
1397 in-person workshops. We thus recruited teachers to facilitate workshops on our behalf, and conducted training sessions
1398 to ensure that they understood their role in the workshops. Many of the limitations mentioned in Section 5.2 emerged
1399
1400
1401
1402
1403
1404

1405 from this mode of carrying out the research and the constraints we operated under at the time. Even so, the workshops
1406 delivered great insights which will inform future research in this space. Moreover, our experiences in carrying out
1407 remote workshops will be instructive for those wishing to carry out similar workshops in the future.
1408

1409 In particular, we discovered that children had a heightened awareness of the activities of bad actors online but
1410 also demonstrated excessive vigilance. This, in turn, hampered their ability to distinguish between dark patterns and
1411 genuine warnings. Importantly, our study identified the need for specific interventions that will help children develop a
1412 more nuanced understanding of online deception and communicate coping mechanisms to recover if they are indeed
1413 deceived by these.
1414

1415 ACKNOWLEDGEMENTS

1416 We are indebted to the teachers for facilitating the workshops on our behalf - none of this would have been possible
1417 without them. We are grateful to our child participants for their enthusiasm and delightful responses. We thank Chelsea
1418 Jarvie for being a safeguarder. We thank SPRITE (EP/S035869/1) and REPHRAIN (EP/W032473/1) for funding this
1419 research. Finally, we thank our anonymous reviewers for their thoughtful comments and suggestions to improve this
1420 paper.
1421
1422
1423

1424 REFERENCES

- 1425
- 1426 [1] Mariek MP Vanden Abeele and Antonius J Van Rooij. 2016. OR-02: Fear Of Missing Out (FOMO) as a predictor of problematic social media use
1427 among teenagers. *Journal of Behavioral Addictions* 5, S1 (2016), 4–5.
 - 1428 [2] Arwa A Al Shamsi. 2019. Effectiveness of cyber security awareness program for young children: A case study in UAE. *Int. J. Inf. Technol. Lang. Stud*
1429 3, 2 (2019), 8–29.
 - 1430 [3] Zahra Ashktorab and Jessica Vitak. 2016. Designing cyberbullying mitigation and prevention solutions through participatory design with teenagers.
1431 In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery, San Jose California, 3895–3905.
1432 <https://doi.org/10.1145/2858036.2858548>.
 - 1433 [4] David Barnard-Wills. 2012. E-safety education: Young people, surveillance and responsibility. *Criminology & Criminal Justice* 12, 3 (2012), 239–255.
1434 <https://doi.org/10.1177/174889581143295>.
 - 1435 [5] Kerstin Bongard-Blanchy, Arianna Rossi, Salvador Rivas, Sophie Doublet, Vincent Koenig, and Gabriele Lenzini. 2021. I am Definitely Manipulated,
1436 even When i am Aware of it. It's Ridiculous! - Dark Patterns from the End-User Perspective. In *Proceedings of the 2021 ACM Designing Interactive*
1437 *Systems Conference: Nowhere and Everywhere (DIS)*. Association for Computing Machinery, Inc, Online, 763–776. <https://doi.org/10.1145/3461778.3462086>.
 - 1438 [6] Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher. 2016. Tales from the Dark Side: Privacy Dark Strategies and
1439 Privacy Dark Patterns. *Proc. Priv. Enhancing Technol.* 2016, 4 (2016), 237–254. <http://doi.org/10.1515/popets-2016-0038>.
 - 1440 [7] Gary L. Brase, Eugene Y. Vasserman, and William Hsu. 2017. Do Different Mental Models Influence Cybersecurity Behavior? Evaluations via
1441 Statistical Reasoning Performance. *Frontiers in Psychology* 8, NOV (11 2017), 1929. <https://doi.org/10.3389/FPSYG.2017.01929>.
 - 1442 [8] Virginia Braun and Victoria Clarke. 2006. Thematic analysis revised. *Qual. Res. Psychol* 3, 2 (2006), 77–101. <http://dx.doi.org/10.1191/1478088706qp063oa>.
 - 1443 [9] Virginia Braun and Victoria Clarke. 2023. Toward good practice in thematic analysis: Avoiding common problems and be(com)ing a knowing
1444 researcher. *International Journal of Transgender Health* 24, 1 (2023), 1–6. [10.1080/26895269.2022.2129597](https://doi.org/10.1080/26895269.2022.2129597).
 - 1445 [10] Harry Brignull. 2020. Types Of Dark Pattern. Accessed 10/11/23 <https://darkpatterns.org/types-of-dark-pattern.html>.
 - 1446 [11] Harry Brignull. 2023. *Deceptive Patterns. Exposing The Tricks Tech Companies Use To Control You*. Testimonium Ltd, UK.
 - 1447 [12] Jessica E. Brodsky, Arshia K. Lodhi, Kasey L. Powers, Fran C. Blumberg, and Patricia J. Brooks. 2021. "It's just everywhere now": Middle-school and
1448 college students' mental models of the Internet. *Human Behavior and Emerging Technologies* 3, 4 (10 2021), 495–511. <https://doi.org/10.1002/hbe2.281>.
 - 1449 [13] Otieno C, Spada H, and Renkl A. 2013. Effects of news frames on perceived risk, emotions, and learning. *PLoS One* 8, 11 (2013), e79696.
1450 [10.1371/journal.pone.0079696](https://doi.org/10.1371/journal.pone.0079696).
 - 1451 [14] L. Jean Camp. 2009. Mental models of privacy and security. *IEEE Technology and Society Magazine* 28, 3 (2009), 37–46. <https://doi.org/10.1109/MTS.2009.934142>.
 - 1452 [15] Victoria Clarke and Virginia Braun. 2013. Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning. *The*
1453 *Psychologist* 26, 2 (2013), 120–123. <https://www.bps.org.uk/psychologist/methods-teaching-thematic-analysis>.
 - 1454 [16] Richard K Coll. 2006. The role of models, mental models and analogies in chemistry teaching. In *Metaphor and analogy in science education*, Peter J.
1455 Aubbesson, Allan G. Harrison, and Stephen M. Ritchie (Eds.). Springer, The Netherlands, 65–77.
1456

- 1457 [17] Gregory Conti and Edward Sobiesk. 2010. Malicious Interface Design: Exploiting the User. In *Proceedings of the 19th International Conference*
1458 *on World Wide Web* (Raleigh, North Carolina, USA) (*WWW '10*). Association for Computing Machinery, New York, NY, USA, 271–280. <https://doi.org/10.1145/1772690.1772719>
1459
- 1460 [18] Dan Cooper, Sam Jungyun Choi, Diane Valat, and Anna Oberschelp de Meneses. 2023. The EU Stance on Dark Patterns. <https://www.insideprivacy.com/eu-data-protection/the-eu-stance-on-dark-patterns/> 31 January.
1461
- 1462 [19] Otávio de Paula Albuquerque, Marcelo Fantinato, Judith Kelner, and Anna Priscilla de Albuquerque. 2020. Privacy in smart toys: Risks and proposed
1463 solutions. *Electronic Commerce Research and Applications* 39 (2020), 100922. <https://doi.org/10.1016/j.elerap.2019.100922>.
1464
- 1465 [20] Pearl Denham. 1993. Nine- to fourteen-year-old children’s conception of computers using drawings. *Behaviour and Information Technology* 12, 6
1466 (1993), 346–358. <https://doi.org/10.1080/01449299308924399>.
1467
- 1468 [21] Linda Di Geronimo, Larissa Braz, Enrico Fregnan, Fabio Palomba, and Alberto Bacchelli. 2020. UI Dark Patterns and Where to Find Them: A Study
1469 on Mobile Applications and User Perception. In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing
1470 Machinery, Virtual, Paper 473. <https://doi.org/10.1145/3313831.3376600>.
1471
- 1472 [22] Stuart Dredge. 2019. All you need to know about Roblox. [https://www.theguardian.com/games/2019/sep/28/roblox-guide-children-gaming-
1473 platform-developer-minecraft-fortnite](https://www.theguardian.com/games/2019/sep/28/roblox-guide-children-gaming-platform-developer-minecraft-fortnite).
1474
- 1475 [23] Stuart Dredge. 2022. Ofcom study explores children’s use of TikTok and YouTube. Accessed 10/11/2023 [https://musically.com/2022/03/31/ofcom-
1476 study-explores-childrens-use-of-tiktok-and-youtube/](https://musically.com/2022/03/31/ofcom-study-explores-childrens-use-of-tiktok-and-youtube/).
1477
- 1478 [24] Martha Driessnack. 2005. Children’s drawings as facilitators of communication: a meta-analysis. *Journal of Pediatric Nursing* 20, 6 (2005), 415–423.
1479 <https://doi.org/10.1016/j.pedn.2005.03.011>.
1480
- 1481 [25] Education Scotland. 2023. Curriculum for Excellence. Curriculum for Excellence documents. Experiences and Outcomes. [https://education.gov.
1482 scot/curriculum-for-excellence/curriculum-for-excellence-documents/experiences-and-outcomes/](https://education.gov.scot/curriculum-for-excellence/curriculum-for-excellence-documents/experiences-and-outcomes/).
1483
- 1484 [26] Susan Edwards, Andrea Nolan, Michael Henderson, Helen Skouteris, Ana Mantilla, Pamela Lambert, and Jo Bird. 2020. Developing a measure
1485 to understand young children’s Internet cognition and cyber-safety awareness: a pilot test. In *Digital Play and Technologies in the Early Years*.
1486 Routledge, London, UK, 100–114.
1487
- 1488 [27] Montserrat Fargas-Malet, Dominic McSherry, Emma Larkin, and Clive Robinson. 2010. Research with children: Methodological issues and innovative
1489 techniques. *Journal of Early Childhood Research* 8, 2 (2010), 175–192. <https://doi.org/10.1177/1476718X09345412>.
1490
- 1491 [28] Valerie S Folkes. 1988. The availability heuristic and perceived risk. *Journal of Consumer Research* 15, 1 (1988), 13–23. <https://doi.org/10.1086/209141>.
1492
- 1493 [29] Kelsey R. Fulton, Rebecca Gelles, Alexandra McKay, Yasmin Abdi, Richard Roberts, and Michelle L. Mazurek. 2019. The Effect of Entertainment
1494 Media on Mental Models of Computer Security. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. USENIX Association, Santa
1495 Clara, CA, 79–95. <https://www.usenix.org/conference/soups2019/presentation/fulton>.
1496
- 1497 [30] Sharon Goldfeld, Elodie O’Connor, Valerie Sung, Gehan Roberts, Melissa Wake, Sue West, and Harriet Hiscock. 2022. Potential indirect impacts of
1498 the COVID-19 pandemic on children: a narrative review using a community child health lens. *Medical Journal of Australia* 216, 7 (2022), 364–372.
1499 <https://doi.org/10.5694/mja2.51368>.
1500
- 1501 [31] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. The Dark (Patterns) Side of UX Design. In *Proceedings of the*
1502 *2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (*CHI '18*). Association for Computing Machinery, New York,
1503 NY, USA, 1–14. <https://doi.org/10.1145/3173574.3174108> <https://doi.org/10.1145/3173574.3174108>.
1504
- 1505 [32] Colin M Gray, Cristiana Santos, Nataliia Bielova, Michael Toth, and Damian Clifford. 2021. Dark patterns and the legal requirements of consent
1506 banners: An interaction criticism perspective. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Virtual,
1507 1–18. <https://doi.org/10.1145/3411764.3445779>.
1508
- 1509 [33] Saul Greenberg, Sebastian Boring, Jo Vermeulen, and Jakub Dostal. 2014. Dark patterns in proxemic interactions: a critical perspective. In *Proceedings*
1510 *of the 2014 Conference on Designing interactive systems*. ACM, Vancouver, Canada, 523–532. <https://doi.org/10.1145/2598510.2598541>.
1511
- 1512 [34] Zhen Guo, Jin-Hee Cho, Ray Chen, Srijan Sengupta, Michin Hong, and Tanushree Mitra. 2020. Online social deception and its countermeasures: A
1513 survey. *IEEE Access* 9 (2020), 1770–1806. <https://doi.org/10.1109/ACCESS.2020.3047337>.
1514
- 1515 [35] Information Commissioner. 2022. The Children’s code design guidance. [https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/
1516 designing-products-that-protect-privacy/childrens-code-design-guidance/](https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/designing-products-that-protect-privacy/childrens-code-design-guidance/).
1517
- 1518 [36] Internet Watch Foundation. 2023. ‘Pivotal moment’ as Online Safety Act gains Royal Assent. [https://www.iwf.org.uk/news-media/news/pivotal-
1519 moment-as-online-safety-act-gains-royal-assent/](https://www.iwf.org.uk/news-media/news/pivotal-moment-as-online-safety-act-gains-royal-assent/).
1520
- 1521 [37] Daniel Kahneman. 2012. *Thinking, Fast and Slow*. Farrar, Straus and Girou, New York.
1522
- 1523 [38] Christie Kodama, Beth St. Jean, Mega Subramaniam, and Natalie Greene Taylor. 2017. There’s a creepy guy on the other end at Google!: engaging
1524 middle school students in a drawing activity to elicit their mental models of Google. *Information Retrieval Journal* 20, 5 (10 2017), 403–432.
1525 <https://doi.org/10.1007/s10791-017-9306-x>.
1526
- 1527 [39] Monica Kowalczyk, Johanna T. Gunawan, David Choffnes, Daniel J Dubois, Woodrow Hartzog, and Christo Wilson. 2023. Understanding Dark
1528 Patterns in Home IoT Devices. In *CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany)
1529 (*CHI '23*). ACM, New York, NY, USA, Article 179, 27 pages. <https://doi.org/10.1145/3544548.3581432> <https://doi.org/10.1145/3544548.3581432>.
1530
- 1531 [40] Maria Lamond, Karen Renaud, Lara Wood, and Suzanne Prior. 2022. SOK: young children’s cybersecurity knowledge, skills & practice: a
1532 systematic literature review. In *Proceedings of the 2022 European Symposium on Usable Security*. ACM, Karlsruhe, Germany, 14–27. <https://doi.org/10.1145/3549015.3554207>.
1533

- 1509 [41] Elmer Lastdrager, Inés Carvajal Gallardo, Pieter Hartel, and Marianne Junger. 2017. How Effective is { Anti-Phishing } Training for Children?. In
1510 *Thirteenth symposium on Usable Privacy and Security (SOUPS 2017)*. USENIX, Santa Clara, CA, 229–239.
- 1511 [42] Isabelle Lee. 2021. A new cryptocurrency called worldcoin wants to scan 1 billion people's iris by 2023 to speed up digital currency adoption. Accessed
1512 10/11/23 <https://markets.businessinsider.com/news/currencies/worldcoin-orb-scan-eyes-iris-sam-altman-y-combinator-cryptocurrency-2021-10>.
- 1513 [43] Pierpaolo Limone and Giusi Antonia Toto. 2021. Psychological and emotional effects of Digital Technology on Children in Covid-19 Pandemic.
1514 *Brain Sciences* 11, 9 (2021), 1126. <https://doi.org/10.3390/brainsci11091126>.
- 1515 [44] Sonia Livingstone, Giovanna Mascheroni, and Mariya Stoilova. 2021. The outcomes of gaining digital skills for young people's lives and wellbeing:
1516 A systematic evidence review. *New Media & Society* 25, 5 (2021), 14614448211043189. <https://doi.org/10.1177/14614448211043189>.
- 1517 [45] Jamie Luguri and Lior Jacob Strahilevitz. 2021. Shining a light on dark patterns. *Journal of Legal Analysis* 13, 1 (2021), 43–109. <https://doi.org/10.1093/jla/laaa006>.
- 1518 [46] Sheri Madigan, Rachel Eirich, Paolo Pador, Brae Anne McArthur, and Ross D Neville. 2022. Assessment of Changes in Child and Adolescent
1519 Screen Time During the COVID-19 Pandemic: A Systematic Review and Meta-analysis. *JAMA Pediatrics* 176, 12 (2022), 1188–1198. <https://doi.org/10.1001/jamapediatrics.2022.4116>.
- 1520 [47] Ana Maria Marhan, Mihai Ioan Micle, Camelia Popa, and Georgeta Preda. 2012. A review of mental models research in child-computer interaction.
1521 *Procedia - Social and Behavioral Sciences* 33 (2012), 368–372. <https://doi.org/10.1016/j.sbspro.2012.01.145>.
- 1522 [48] Theresa M Marteau, Paul C Fletcher, Marcus R Munafò, and Gareth J Hollands. 2021. Beyond choice architecture: advancing the science of changing
1523 behaviour at scale. *BMC Public Health* 21, 1 (2021), 1–7. <https://doi.org/10.1186/s12889-021-11382-8>.
- 1524 [49] Alice E Marwick and danah boyd. 2014. Networked privacy: How teenagers negotiate context in social media. *New Media & Society* 16, 7 (2014),
1525 1051–1067. <https://doi.org/10.1177/1461444814543995>.
- 1526 [50] Arunesh Mathur, Gunes Acar, Michael J Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark patterns
1527 at scale: Findings from a crawl of 11K shopping websites. In *Proceedings of the ACM Conference on Human-Computer Interaction*, Vol. 3. ACM, New
1528 York, NY, USA, 1–32. <https://doi.org/10.1145/3359183>.
- 1529 [51] Thomas Mildner. 2020. Thomas' Dark Pattern Cheatsheet. Accessed 10/11/23 <https://thomasmildner.me/darkpatterns.html>.
- 1530 [52] Ann Minckler. 2006. *Middle school children online: Comparing parent awareness and supervision of students' behaviors*. Ph.D. Dissertation. University
1531 of Montana.
- 1532 [53] Benjamin Morrison, Cigdem Sengul, Mark Springett, Jacqui Taylor, and Karen Renaud. 2021. WHITE PAPER: Mental Models of Dark Patterns.
1533 SPRITE White Paper https://spritehub.org/wp-content/uploads/2021/12/SPRITE_Lit_Review8.pdf.
- 1534 [54] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. 2020. Dark Patterns after the GDPR: Scraping Consent Pop-Ups and
1535 Demonstrating Their Influence. In *Proceedings CHI (Honolulu, HI, USA) (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13.
<https://doi.org/10.1145/3313831.3376321>.
- 1536 [55] Maggie Oates, Yama Ahmadullah, Abigail Marsh, Chelse Swoopes, Shikun Zhang, Rebecca Balebako, and Lorrie Faith Cranor. 2018. Turtles, Locks,
1537 and Bathrooms: Understanding Mental Models of Privacy Through Illustration. *Proceedings on Privacy Enhancing Technologies* 2018, 4 (10 2018),
1538 5–32. <https://doi.org/10.1515/popets-2018-0029>.
- 1539 [56] Ofcom. 2020. Parents' rising concern over children online. <https://www.ofcom.org.uk/about-ofcom/latest/media/media-releases/2020/rising-concern-over-children-online>.
- 1540 [57] Ofcom. 2021. Children's News Consumption Survey 2021. https://www.ofcom.org.uk/_data/assets/excel_doc/0028/219880/childrens-news-consumption-survey-data-tables.xlsx.
- 1541 [58] Ofcom. 2022. Online Nation 2022 report. https://www.ofcom.org.uk/_data/assets/pdf_file/0023/238361/online-nation-2022-report.pdf Accessed 17
1542 July 2023.
- 1543 [59] Nils Pancratz and Ira Diethelm. 2020. "Draw us how smartphones, video gaming consoles, and robotic vacuum cleaners look like from the inside":
1544 Students' conceptions of computing system architecture. In *PervasiveHealth: Pervasive Computing Technologies for Healthcare*. ICST, online, 1–10.
<https://doi.org/10.1145/3421590.3421600>.
- 1545 [60] Gordon Pask and Bernard CE Scott. 1972. Learning strategies and individual competence. *International Journal of Man-Machine Studies* 4, 3 (1972),
1546 217–253. [https://doi.org/10.1016/S0020-7373\(72\)80004-X](https://doi.org/10.1016/S0020-7373(72)80004-X).
- 1547 [61] Suzanne Prior and Karen Renaud. 2022. The impact of financial deprivation on children's cybersecurity knowledge & abilities. *Education and
1548 Information Technologies* 27 (2022), 10563–83. <https://doi.org/10.1007/s10639-022-10908-w>.
- 1549 [62] Suzanne Prior and Karen Renaud. 2023. Who is best placed to support cyber responsabilized UK parents? *Children* 10, 7 (2023), 1130. <https://doi.org/10.3390/children10071130>.
- 1550 [63] Pavol Prokop, Jana Fančovičová, and Sue Dale Tunnicliffe. 2009. The Effect of Type of Instruction on Expression of Children's Knowledge:
1551 How Do Children See the Endocrine and Urinary System? *International Journal of Environmental and Science Education* 4, 1 (1 2009), 75–93.
<http://www.ijese.com/>.
- 1552 [64] Samantha Punch. 2002. Research with children: The same or different from research with adults? *Childhood* 9, 3 (2002), 321–341. <https://doi.org/10.1177/0907568202009003005>.
- 1553 [65] Farzana Quayyum, Daniela S Cruzes, and Letizia Jaccheri. 2021. Cybersecurity awareness for children: A systematic literature review. *International
1554 Journal of Child-Computer Interaction* 30 (2021), 100343. <https://doi.org/10.1016/j.ijcci.2021.100343>.
- 1555
1556
1557
1558
1559
1560

- 1561 [66] Emilee Rader and Rick Wash. 2015. Identifying patterns in informal sources of security information. *Journal of Cybersecurity* 1, 1 (2015), 121–144.
1562 <https://doi.org/10.1093/cybsec/tyv008>.
- 1563 [67] Karen Renaud and Marc Dupuis. 2019. Cyber security fear appeals: Unexpectedly complicated. In *Proceedings of the New Security Paradigms*
1564 *Workshop*. ACM, Santa Cruz, USA, 42–56. <https://doi.org/10.1145/3368860.3368864>.
- 1565 [68] Karen Renaud and Suzanne Prior. 2021. The “three M’s” counter-measures to children’s risky online behaviors: mentor, mitigate and monitor.
1566 *Information & Computer Security* 29, 3 (2021), 526–557. <https://doi.org/10.1108/ICS-07-2020-0115>.
- 1567 [69] Karen Renaud and Verena Zimmermann. 2019. Nudging folks towards stronger password choices: providing certainty is the key. *Behavioural Public*
1568 *Policy* 3, 2 (2019), 228–258. <https://doi.org/10.1017/bpp.2018.3>.
- 1569 [70] J Ritchie and L Spencer. 1994. Qualitative data analysis for applied policy research. In *Analyzing Qualitative Data*, Alan Bryman and Bob Burgess
1570 (Eds.). Taylor & Francis, London, New York, Chapter 9, 173–194. <https://doi.org/10.4324/9780203413081>.
- 1571 [71] Jane Ritchie, Liz Spencer, and William O’Connor. 2003. Carrying out qualitative analysis. In *Qualitative research practice: A guide for social science*
1572 *students and researchers*, Jane Ritchie and Jane Lewis (Eds.). Sage, London, Chapter 9, 219–62.
- 1573 [72] Laura Rook. 2013. Mental models: A robust definition. *The Learning Organization* 20, 1 (2013), 38–47. <https://doi.org/10.1108/09696471311288519>.
- 1574 [73] William B Rouse and Nancy M Morris. 1986. On looking into the black box: Prospects and limits in the search for mental models. *Psychological*
1575 *Bulletin* 100, 3 (1986), 349–363. <https://doi.org/10.1037/0033-2909.100.3.349>.
- 1576 [74] Anna L Rowe and Nancy J Cooke. 1995. Measuring mental models: Choosing the right tools for the job. *Human Resource Development Quarterly* 6, 3
1577 (1995), 243–255. <https://doi.org/10.1002/hrdq.3920060303>.
- 1578 [75] Anna L. Rowe, Nancy J. Cooke, Kelly J. Neville, and Chris W. Schacherer. 1992. Mental models of mental models: a comparison of mental model
1579 measurement techniques. *Proceedings of the Human Factors Society* 2 (1992), 1195–1199. <https://doi.org/10.1177/154193129203601603>.
- 1580 [76] Eliza Rybska, Sue Tunnicliffe, and Zofia Chyleńska. 2014. Young children’s ideas about snail internal anatomy. *Journal of Baltic Science Education* 13
1581 (12 2014), 828–838. <https://doi.org/10.33225/jbse/14.13.828>.
- 1582 [77] Statista. 2021. Level of difficulty identifying whether a news story on social media is true among children in the United Kingdom (UK) as of March
1583 2023. <https://www.statista.com/statistics/1268672/children-identifying-trustworthy-news-online-united-kingdom-uk/>.
- 1584 [78] Timothy Summers, Kalle J Lyytinen, Tony Lingham, and Eugene A Pierce. 2013. How hackers think: A study of cybersecurity experts and
1585 their mental models. In *Third Annual International Conference on Engaged Management Scholarship*. EDBAC, Atlanta, Georgia, Paper 3.3. <http://dx.doi.org/10.2139/ssrn.2326634>.
- 1586 [79] Richard H Thaler and Cass R Sunstein. 2007. *Nudge: Improving decisions about health, wealth, and happiness*. HeinOnline, USA.
- 1587 [80] Sue Dale Tunnicliffe and Michael J Reiss. 1999. Building a model of the environment: how do children see animals? *Journal of Biological Education*
1588 33, 3 (1999), 142–148. <https://doi.org/10.1080/00219266.1999.9655654>.
- 1589 [81] Sue Dale Tunnicliffe and Michael J. Reiss. 2000. Building a model of the environment: how do children see plants? *Journal of Biological Education* 34,
1590 4 (2000), 172–177. <https://doi.org/10.1080/00219266.2000.9655714>.
- 1591 [82] James Turland, Lynne Coventry, Debora Jeske, Pam Briggs, and Aad van Moorsel. 2015. Nudging towards security: Developing an application
1592 for wireless network selection for android phones. In *Proceedings of the 2015 British HCI conference*. ACM, Lincoln, USA, 193–201. <https://doi.org/10.1145/2783446.2783588>.
- 1593 [83] Amos Tversky and Daniel Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185, 4157 (1974), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>.
- 1594 [84] UNICEF. 2021. Policy guidance on AI for children. <https://www.unicef.org/globalinsight/reports/policy-guidance-ai-children>.
- 1595 [85] US Court Case. 2023. COMPLAINT FOR INJUNCTIVE AND OTHER RELIEF. <https://oag.ca.gov/system/files/attachments/press-docs/>.
- 1596 [86] Ellen Van Gool, Joris Van Ouytsel, Koen Ponnet, and Michel Walrave. 2015. To share or not to share? Adolescents’ self-disclosure about peer
1597 relationships on Facebook: An application of the Prototype Willingness Model. *Computers in Human Behavior* 44 (2015), 230–239. <https://doi.org/10.1016/j.chb.2014.11.036>.
- 1598 [87] Mariek Vanden Abeele, Scott W Campbell, Steven Eggermont, and Keith Roe. 2014. Sexting, mobile porn use, and peer group dynamics: Boys’ and
1599 girls’ self-perceived popularity, need for popularity, and perceived peer pressure. *Media Psychology* 17, 1 (2014), 6–33.
- 1600 [88] Rojin Vishkaie. 2021. Companion toys for children: using drawings to probe happiness. *Interactions* 28, 4 (2021), 39–43. <https://doi.org/10.1145/3466166>.
- 1601 [89] Joyce Vissenberg, Leen d’Haenens, and Sonia Livingstone. 2022. Digital Literacy and Online Resilience as Facilitators of Young People’s Well-Being?
1602 *European Psychologist* 27, 2 (2022), 76–85. <https://doi.org/10.1027/1016-9040/a000478>.
- 1603 [90] Melanie Volkamer and Karen Renaud. 2013. Mental models—general introduction and review of their application to human-centred security. In
1604 *Number Theory and Cryptography*, Marc Fischlin and Stefan Katzenbeisser (Eds.). Springer, Berlin, Germany, 255–280. https://doi.org/10.1007/978-3-642-42001-6_18.
- 1605 [91] Rick Wash. 2010. Folk models of home computer security. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*. ACM, Redmond,
1606 USA, 1–16. <https://doi.org/10.1145/1837110.1837125>.
- 1607 [92] Rick Wash and Emilee Rader. 2011. Influencing mental models of security: a research agenda. In *Proceedings of the 2011 New Security Paradigms*
1608 *Workshop*. ACM, California, USA, September 12 - 15, 57–66. <https://doi.org/10.1145/2073276.2073283>.
- 1609
1610
1611
1612

Table 11. Transcription themes, Sub-Themes, Codes and Examples

Theme	Sub-Theme	Codes	Examples from Transcripts
Dark Pattern Perceptions	Leaked Data via user interaction or snooping	Web account; Personal details; phone number; credit card;	“So when you click the button for free Roblox, it says email address, required, password required and bank details required. So you can't claim the free Robux without doing anything” (WS4S2C2)
	Nefarious Actor	Scammer; Guy behind the computer; Hacker capabilities	“Probably like someone who a scammer and tries to sell things and like take money off people” (WS5S2C2); “but the guy behind the computer that got the information would be ha ha ha”. (WS6S1C3) “I think they might be smart and find a way to kind of get it without you having to enter it like without having to enter in your details” (WS6S4C1)
	Action	Identified correct scenario; things work as normal	“So he clicks on the open your camera and then he ends up going to put in your details and then he sends a message saying get hacked LOL” (WS4S1C1) “You're going to get hacked because if you read the website name. The website is instagram.con ” (WS3S4C2) (about back to safety button) I think that would probably take you back to your website or close your app. (WS4S4C1)
	Consequences	Nothing bad; offline harm; leaked data shared with others; or used to scam/spam the victim/friends/others; bank account hacked & financial loss; web account(s)/device hacked or infected with a virus	“The back to safety button would also give you a virus. Either way you are going to get a virus” (WS4S4C4) “See my face and then will come to my house. The camera can see your school uniform. Then they will see my badge and then they can come to school” (WS6S1C2) “I think it gets your details and streamed online” (WS3S4C2) “It takes money from your bank account and you get scammed” (WS1S3C2)
Online Behaviours	Cautious Practices	Leave game; do not input details; close/refresh/delete app; report incident or 'dodgy' site; use fake email;	“I would just leave the game” (WS6S1C1) “I wouldn't put my details” (WS5S2C2) “And the best way to prevent this by reporting in the person or contacting Roblox” (WS3S2C4)
	Risky Practices	Skip ToS; ignore age notice; brand bias;	“One of these games, if you really wanna play. But then the only reason get in is if you accept all terms of service. I can't be bothered. Click the accept button and then yeah, just click not now” (WS6S1C4)
Security Knowledge	Parental Influence	Ask mom; mom's experience	“And I always ask my mom if I can get games” (WS6S1C5) “Mum was hacked once on email” (WS7S1C1)
	Schools & Teachers	School teaches online safety; School blocks harmful content; ask teacher	“We have restrictions on our iPads. So then we can't get into every website comes up, some of them blocked” (WS6S4C3) “Don't trust things like that. My teachers always said that as well” (WS7S2C1)
	Familiarity with Terminology	No knowledge; knows bad links, browser lock, remote PC; malware/virus/trojans/dark web; own/peer experience	“He could not click robux because it would download malware on your computer” (WS3S2C4) “That when you click the link and you go to a page and then the page says enter your bank details. And they are all required like security code and sort code and they will [say] We won't use your card for any payments. The link is called Twitter.com dot scam dot Mexico” (WS4S2C3)

Table 12. : Framework Analysis Coding

Basic Reading	Basic Reading Definition
(1) input-only data access	Mirrors our scenario back to us; user input (either keyboard or camera); not intrusive; user gives access Example: tit-for-tat device access or access with consent e.g., email address check (without password) or biometrics authentication.
(2) deception-based data / device access & forced downloads	Starts to imagine more than what is presented; information is being tracked or collected via other invasive means without user input, and beyond necessary info, e.g.: access to location services that contributes to tracking; camera access for the full photo of the user; forced video viewing; access to an application store and/or request for or installation of (unnecessary applications); or downloads of software that may induce further tracking or installation of malware.
(3) Leaked account credentials	Begins to identify what is being lost in terms of security/privacy. Full account details (non-biometric) e.g.: the collection of account / login credentials (i.e. username/email and password) to the correct account for the website being visited or to other accounts such as YouTube, Google account.
(4) Leaked sensitive PII (personally identifiable information)	Privacy and security danger; hacking of other sensitive/personally identifiable data Collection of sensitive or personally identifiable information (PII) , e.g. bank details and PII such as an address, postcode, date of birth, age, or phone number.

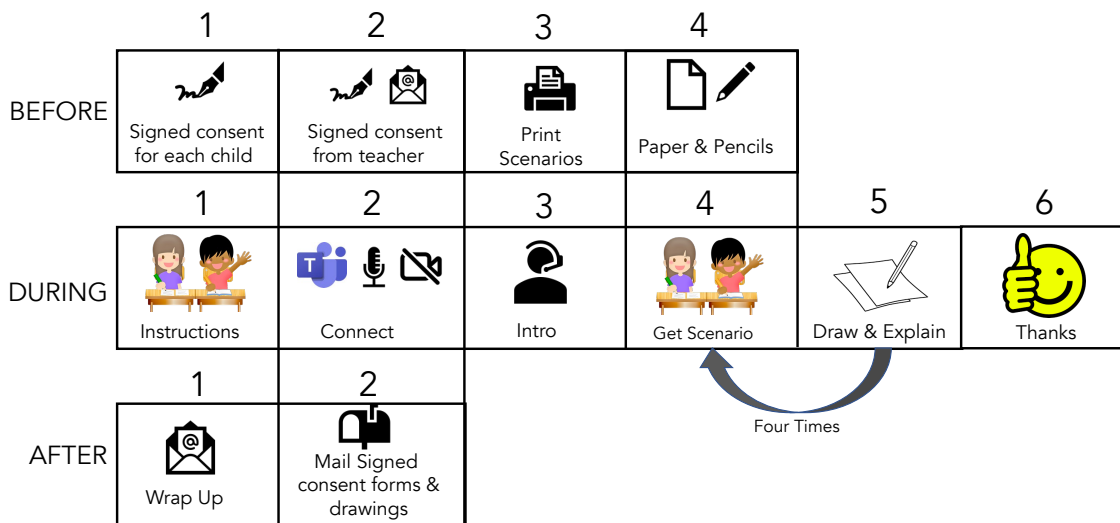


Fig. 7. Bird's Eye View of Workshop

A ETHICS

We carefully designed the workshop with the participant children's safety and privacy being the most important consideration. Figure 7 provides a bird's eye view of the workshop.

- (1) The parents sign consent for their children to participate in the workshops, and give the researchers licence to use the drawings in future publications.
- (2) A teacher facilitates the workshop, and signs a consent form to do this.

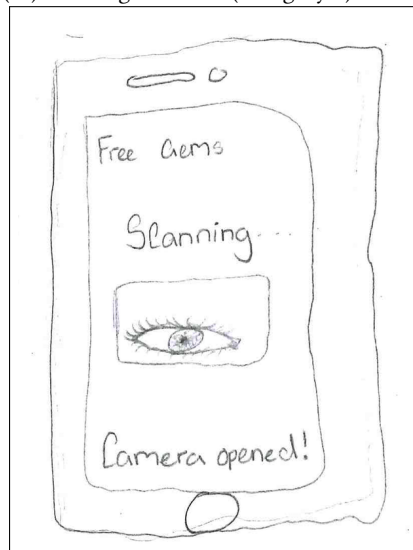
- 1717 (3) We use Microsoft Teams, as the most secure option
- 1718 (4) A safeguarder will be in the Teams room – not participating but having the authority to call a halt if anything is
- 1719 considered to be unsafe for the participants.
- 1720 (5) The camera is not switched on to preserve the anonymity of the children.
- 1721 (6) The researchers record the discussions and the audio files will be stored on the University of Strathclyde’s
- 1722 secure servers.
- 1723 (7) The lead researcher ensures that children’s names, if they are mentioned, are removed before transcription
- 1724 takes place.
- 1725 (8) Once transcribed, recordings are destroyed to preserve anonymity.
- 1726 (9) Drawings will be sent to a team member to ensure that no identifying information has inadvertently been
- 1727 included. Any that appear will be removed before drawings are made available for analysis to the research team.
- 1728 (10) Children receive a certificate of participation – these are printed and sent to the school so the teacher can write
- 1729 the children’s names on them (once again to preserve anonymity).
- 1730 (11) The school receives a sum of money so that they can buy equipment for their classroom.
- 1731 (12) Teachers receive a voucher to thank them for their facilitation of the workshop.
- 1732
- 1733
- 1734
- 1735

1736 Strathclyde Ethics Approval #1619; Northumbria #39624; Brunel #32792-MHR-Nov/2021-4711-1; the other institutions
1737 accepted Strathclyde’s Approval. The data was stored on the University of Strathclyde’s secure servers. All consent
1738 forms and drawings are stored in a locked cabinet and will be retained for 10 years as required by the lead author’s
1739 institution.
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768

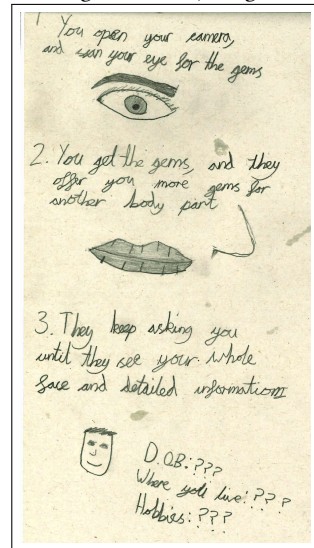
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820

Fig. 8. Scenario 1 Drawings
*

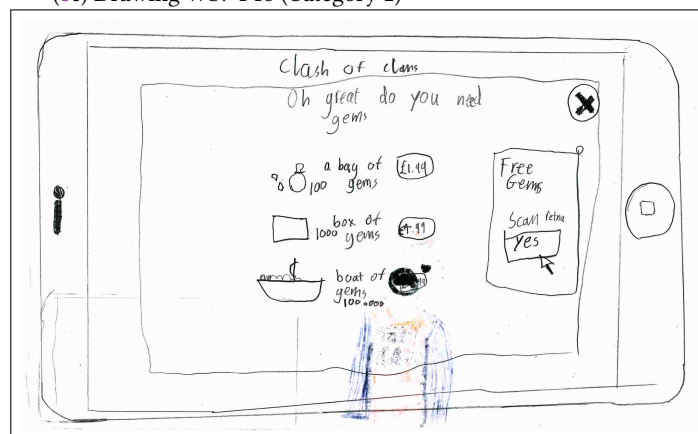
(8a) Drawing WS3-P23 (Category 1)



(8b) Drawing WS6-P12 (Categories 1, 2, and 4.)



(8c) Drawing WS7-P18 (Category 2)



1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872

Fig. 8. Scenario 1 Drawings

(8e) Drawing WS7-P08 (Category 4)

(8d) Drawing WS4-P10 (Category 3)

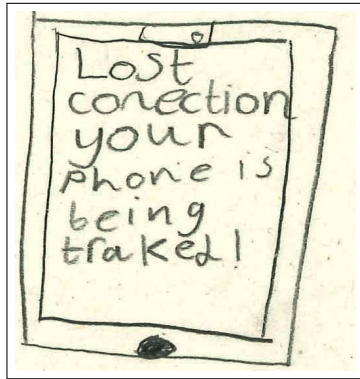


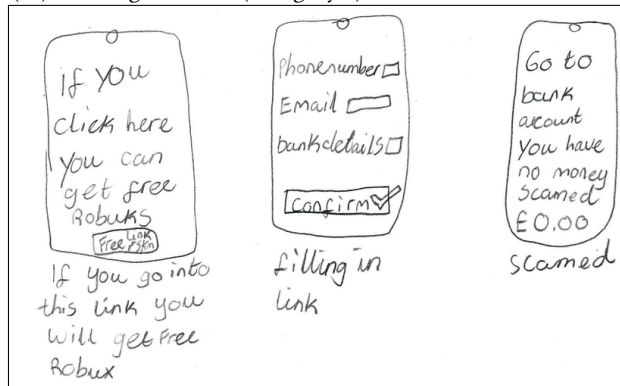
Fig. 9. Scenario 2 Drawings

(9a) Drawing WS5-P17 (Category 2)

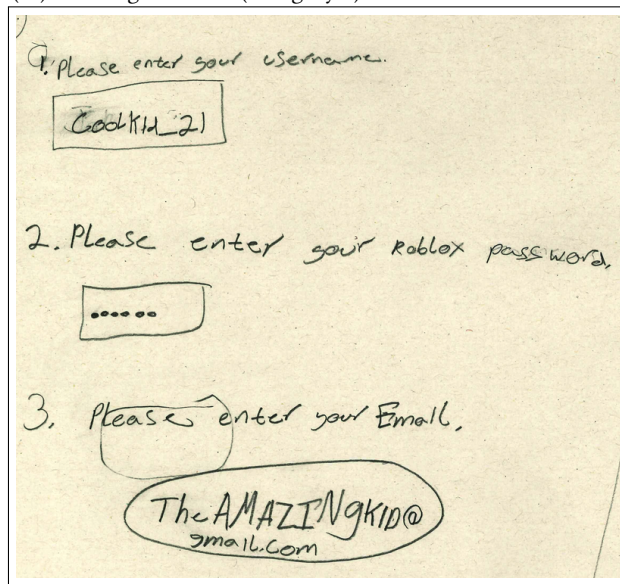


Fig. 9. Scenario 2 Drawings

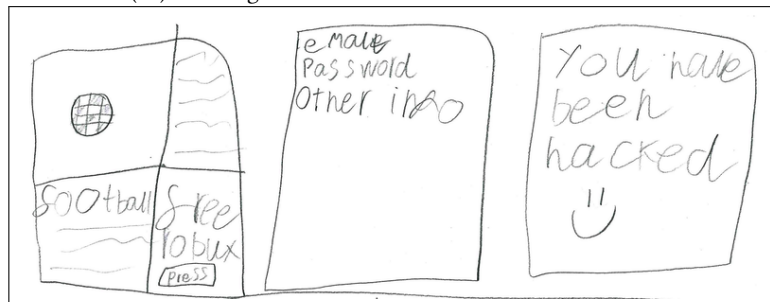
(9b) Drawing WS5-P11 (Category 3)



(9c) Drawing WS6-P08 (Category 3)



(9d) Drawing WS3-P21

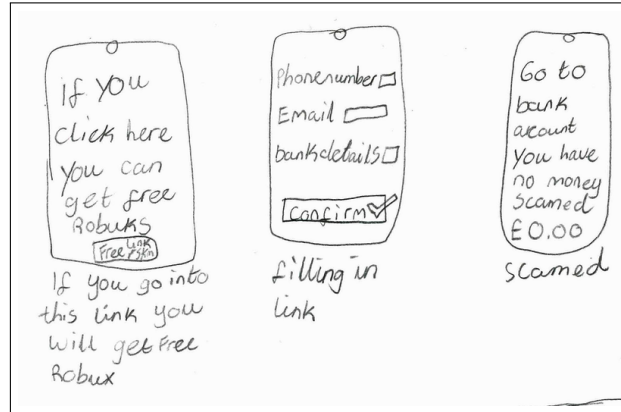


1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924

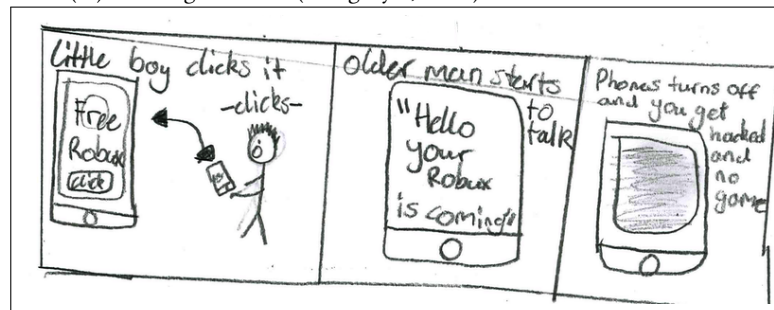
Fig. 9. Scenario 2 Drawings

1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976

(9e) Drawing WS5-P22



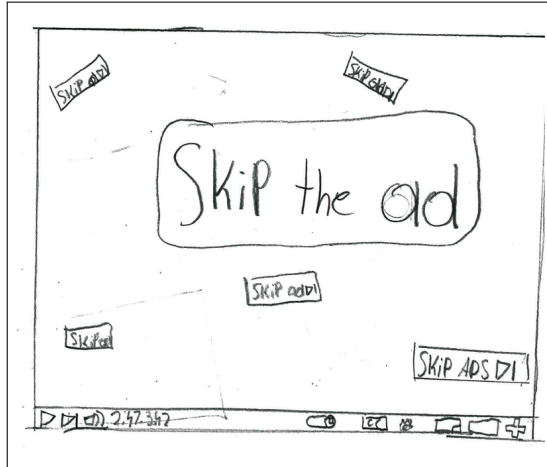
(9f) Drawing WS3-P05 (Category 1, 3 & 4)



1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028

Fig. 10. Scenario 3 Drawings

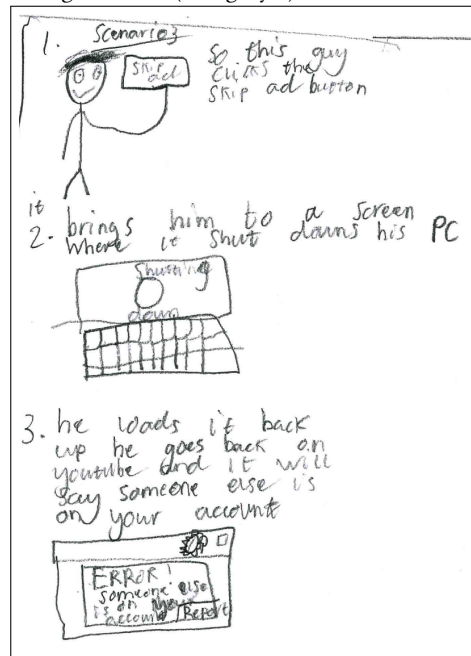
(10a) Drawing WS3-P04 (Category 1)



(10b) Drawing WS5-P05 (Category 2)



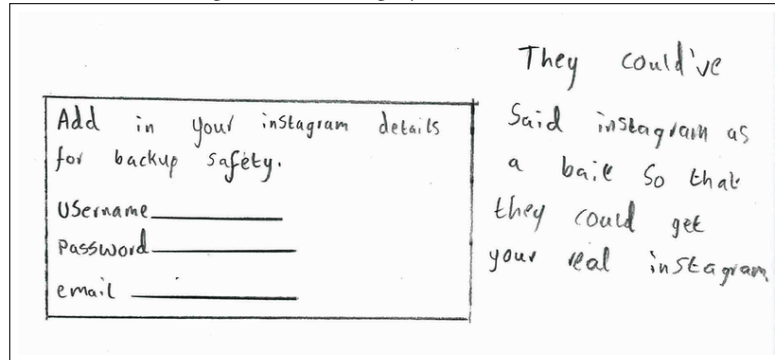
(10c) Drawing WS3-P12 (Category 4)



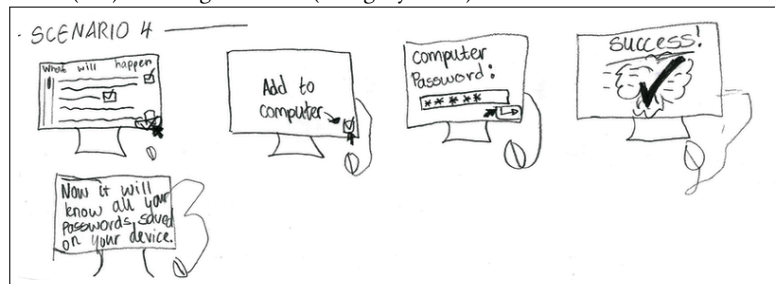
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080

Fig. 11. Scenario 4 Drawings

(11a) Drawing WS7-P09 (Category 1)



(11b) Drawing WS2-P01 (Category 2 & 4)

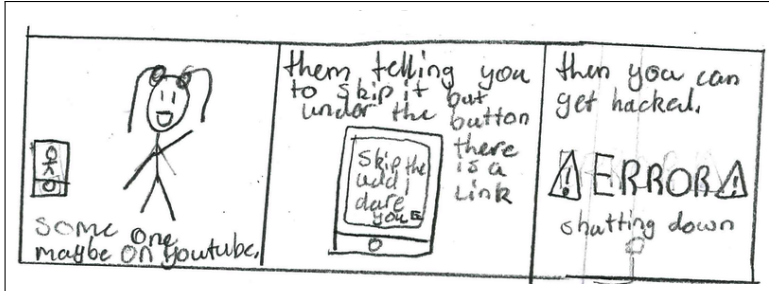


(11c) Drawing WS7-P16 (Category 2 & 3)

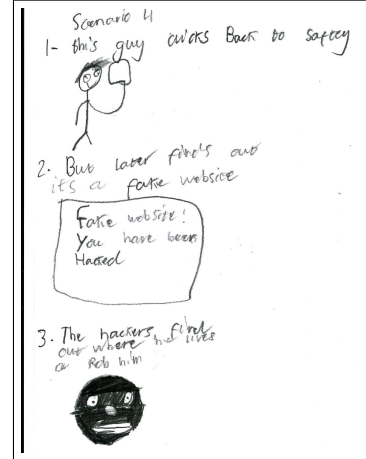


Fig. 12. Discussion Drawings

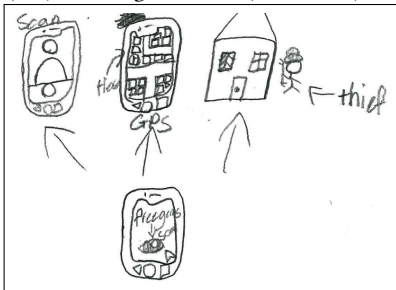
(12a) Drawing WS3-P05 (Scenario 3)



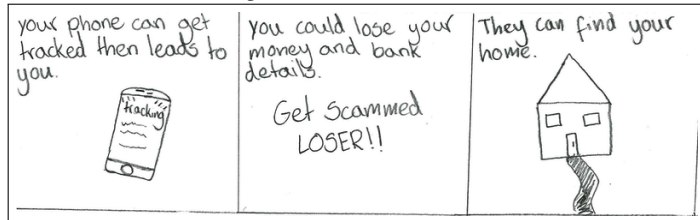
(12b) Drawing WS3-P01 (Scenario 4)



(12c) Drawing WS3-P08 (Scenario 1)



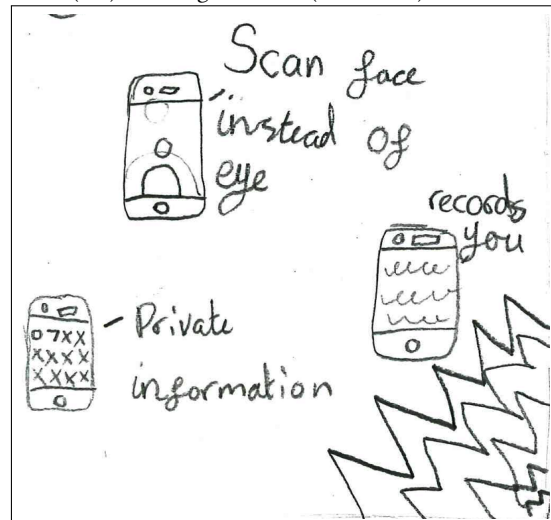
(12d) Drawing WS3-P19 (Scenario 1)



(12e) Drawing WS4-P20 (Scenario 1)



(12f) Drawing WS3-P30 (Scenario 1)



2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132

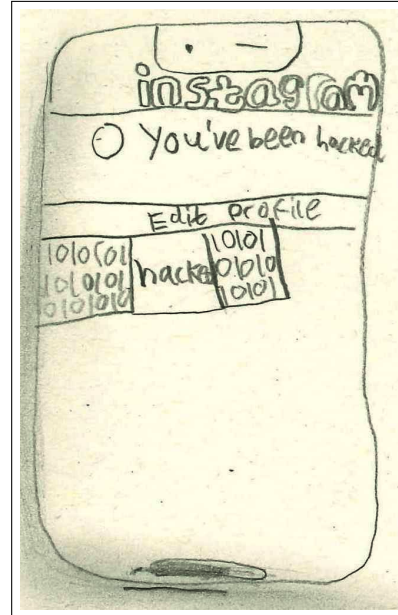
Fig. 12. Discussion Drawings

2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184

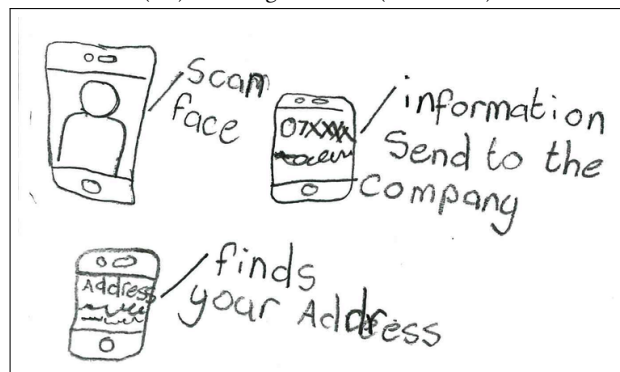
(12g) Drawing WS7-P19 (Scenario 1)



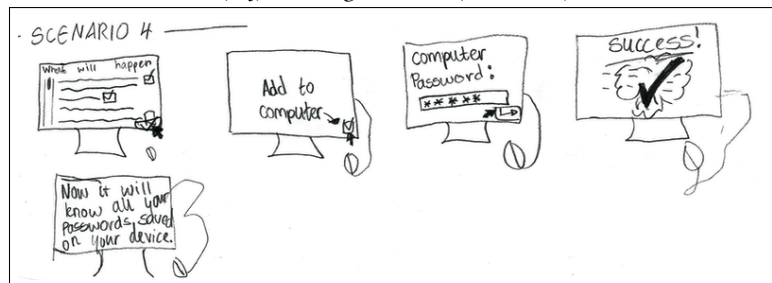
(12h) Drawing WS4-P06 (Scenario 4)



(12i) Drawing WS3-P29 (Scenario 1)



(12j) Drawing WS2-P01 (Scenario 4)



2185 **B AUTHOR STATEMENT**

2186 The ethical design of these workshops was published in a white paper (as required by our funder). However, this was
2187 published before the workshops took place. As such, this paper reports on our findings and the analysis of the drawings
2188 and transcripts, and suggestions for future work based on our new insights. The design and ethical considerations are
2189 included for the sake of completeness here.
2190
2191

2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236