Electronic Commerce Research and Applications 42 (2020) 100984

Contents lists available at ScienceDirect

Electronic Commerce Research and Applications

journal homepage: www.elsevier.com/locate/elerap

Usability heuristics on parental privacy controls for smart toys: From an exploratory map to a confirmatory research

André de Lima Salgado[a,*], Renata Pontin de Mattos Fortes[a], Ricardo Ramos de Oliveira[b], André Pimenta Freire[c]

[a] ICMC, University of São Paulo (USP), São Carlos, Brazil
[b] Federal Institute of Education, Science and Technology of South of Minas Gerais, IFSULDEMINAS, Poços de Caldas, Brazil
[c] Federal University of Lavras, UFLA, Brazil

ARTICLE INFO

Keywords:
Smart toy
Privacy
Usability
Heuristic evaluation
Parental control
Inspection
Review

ABSTRACT

In this paper, we aimed to indicate usability heuristics for the design of parental privacy controls for smart toys. During a snowballing mapping process, we examined 589 candidate studies. Our mapping findings draw from 13 included studies and indicate the heuristics for IT Security Management, proposed by Jaferian et al., as the best to address problems that affect laypeople's interaction with privacy policy tools. With the participation of 14 inspectors, we compared the effectiveness of Nielsen's and the IT Security Management heuristics in heuristic evaluations of a parental privacy control model for smart toys. The results show that the IT Security Management heuristics have better coverage of usability problems (Kruskal–Wallis, p-value ≈ 0.01), which confirms the mapping findings. Future studies can compare these heuristics based on outcomes from test with users as a benchmark. Also, future studies can explore the creation of domain-specific heuristics for parental privacy controls for smart toys.

1. Introduction

Toys are "*any product or material designed or clearly intended for use in play by children under 14 years of age*" (ISO, 2018). They have been present in the daily life of human society for thousands of years; currently, they are part of the life of billions of individuals (Rafferty et al., 2017). Fostering this market, smart toys that listen and interact with children have recently gained popularity (Mahmoud et al., 2018; Valente and Cardenas, 2017; McReynolds et al., 2017). According to a study by Juniper Research, sales of smart toys are expected to grow threefold and exceed $15.5 billion dollars by 2022 (Juniper Research, 2017).

Although traditional toys raise little concern for child's privacy in general, smart toys are able to collect users' contextual data (e.g. location and time) and physical activity (e.g. voice). Such data collection is needed by service providers so that smart toys can learn about users' behavior and provide personalized services (Rafferty et al., 2017; Kaushik et al., 2018). Data sharing may have economic advantages (Acquisti et al., 2016). Yet, it raises important privacy concerns (Rafferty et al., 2017; Kaushik et al., 2018). According to UNICEF (United Nations Children's Fund), one of the major regulatory mechanisms to protect children's privacy online is requiring parental consent prior to the processing of children's personal data. This includes the US law "*Children's Online Privacy Protection Act*" (COPPA) and the General Data Protection Regulation (GDPR), from the European Union. Countries such as South Africa and Spain have similar provisions, and the UK provides recommendations by the Information Commissioner's Office (Children's Commissioner, 2018). Under these circumstances, parents strive to protect children's privacy, and parental privacy controls are seen as a promising approach to solve the problem of undue exposure of children's information by using smart toys (Rafferty et al., 2017).

Parental privacy control is a "*feature in a smart toy for the parents to restrict the content the children can provide to the toy*" (Rafferty et al., 2017, p. 1227). As a privacy control, it aims at reducing the likelihood or the consequences of privacy risks, which may include specification of privacy policies (Garfinkel and Lipford, 2014; Paci et al., 2018; ISO/IEC, 2011). However, usability of privacy policy specification remains one of the main challenges in usable privacy (Garfinkel and Lipford, 2014; Paci et al., 2018; Bertino, 2016; De and Zezschwitz, 2016). Without usable tools, even experts are likely to misconfigure and to leave unwanted vulnerabilities in a smart toy (Sasse and Smith, 2016). More usable privacy tools for laypeople are necessary (Bertino, 2016), as many parents and legal guardians are not necessarily specialists in IT

* Corresponding author.
E-mail address: alsalgado@usp.br (A. de Lima Salgado).

https://doi.org/10.1016/j.elerap.2020.100984
Received 1 May 2019; Received in revised form 21 January 2020; Accepted 14 April 2020
Available online 04 June 2020

and on privacy settings in digital systems.

This study attempts to support the design of more usable parental privacy controls by enhancing its evaluation stage. We focused on one of the most popular usability inspection methods (Lazar et al., 2017; Alonso-Ríos et al., 2018), the Heuristic Evaluation method. Since usability heuristics may emerge as new domains arise (e.g. smart toys), domain-focused heuristics have been proposed to avoid that the usability of interactive systems in new domains be overlooked (Hermawati and Lawson, 2016; de Lima Salgado et al., 2016; de Lima Salgado and Freire, 2014).

In this paper, we sought to systematize existing knowledge on usability heuristics for the domain of lay privacy policy interfaces, and to identify those with the best effectiveness in inspections of parental privacy controls for smart toys. We mapped the literature driven by the question: *What usability heuristics best address usability problems that affect laypeople's interaction with privacy policy interfaces related to smart toys?* Conducting a snowballing procedure (Wohlin, 2014), we examined 589 candidate studies and performed a systematic analysis on 13 included studies. Furthermore, we conducted an empirical case study with 14 usability inspectors to confirm the findings from the literature mapping.

The remaining of this paper is organized as follows. Section 2 presents the background of this study, including terms and definitions and an overview on smart toys and parental privacy controls. In Section 3, we describe the research design of this study, indicating how we have joint the mapping study with the empirical case study to answer our question and reach our goal. Section 4 details the snowballing procedures of the literature mapping, indicating its cumulative outcomes. In Section 5, we present the mapping results and discuss them according to criteria from the literature. We then indicate the most appropriate heuristic set to answer the question, according to the mapping results. Therefore, we describe a case study conducted to confirm the answer. Finally, in Section 7 we sum up the conclusions of this study and indicate important topics for future research usable parental privacy controls for smart toys.

2. Background

2.1. Usable privacy

Usable privacy is the research field aimed at studying the usability of systems that help end-users or administrators to manage data privacy (De and Zezschwitz, 2016; Garfinkel and Lipford, 2014). The ISO/TR 18638 (ISO/TR, 2018) defines *information privacy* as "*rights and obligations of individuals and organizations with respect to the collection, use, retention, disclosure and disposal of personal information*". As indicated, Personally Identifiable Information (PII) is a key term in such definition, and is presented by the ISO/IEC 29100 (ISO/IEC, 2011) as:

> Any information that (a) can be used to identify the Personally Identifiable Information (PII) principal to whom such information relates, or (b) is or might be directly or indirectly linked to a PII principal.

Meanwhile, the ISO 9241-11 (ISO, 2010) defines usability as:

> the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.

The interest on usable privacy had a rapid development during the past two decades (Still, 2016; De and Zezschwitz, 2016; Garfinkel and Lipford, 2014; Cranor and Buchler, 2014). First, both industry and academic literature have seen usability and security (including privacy) only as antagonistic, and users seen as the greatest risk to information security (Still, 2016). Later, because the range of potential threats increased due to the pervasiveness of data (Bertino, 2016), laypeople were often required to make security decisions (Jang-Jaccard and Nepal, 2014) and were seen as the "*greatest hope*" for the area (Still, 2016). Without usable tools, even experts are likely to misconfigure and to leave vulnerabilities in systems (Sasse and Smith, 2016). In such cases, security breaches may be attributed to designers rather than to laypeople (Wash and Zurko, 2017). Nevertheless, even fundamental concepts of laypeople's interaction with privacy tools, as mental models (Oates et al., 2018), remain rare in the literature. The usable privacy field still needs usable tools for laypeople (Bertino, 2016); also, it still needs appropriate usability methods for this domain. In the meantime, governments have been proposing different privacy-related regulations regarding the use of cloud-connected devices, such as smart toys.

To serve as examples, we briefly describe two legislative regulations, comparing them with traditional usability principles, the usability heuristics of Nielsen (Nielsen, 2018). The two legislations are the Children's Online Privacy Protection Rule (COPPA), and the General Data Protection Regulation (GDPR). We chose these regulations because they require parental consent regarding children's data privacy.

In 1998, the United States Congress enacted COPPA (Federal Trade Commission, 2013). By the end of 2012, the United States Congress issued an amended Rule to COPPA, which considered new categories of information (particularly used by connected devices, as geolocation, usernames, child's photos and videos, and persistent Web identifiers) to their definition of personal information[1] COPPA current version is dated from January 17, 2013 (Federal Trade Commission, 2013). In 2016, the European Parliament and its Council created the GDPR (The European Parliament, 2016). Its latest update dates back from April 27th, 2016. The GDPR became applicable in all European Union Member States on May 2018 (EU, 2018, p. 17). By reviewing COPPA and GDPR we can identify usability criteria that impact the design of these technologies based on its similarity with some of the traditional ten usability heuristics from Nielsen. These heuristics are broad usability rules created from sets of known usability problems and their potential solutions (Nielsen, 1994); they are criteria for usability evaluation (ISO, 2010). Examples of relations between regulations and usability criteria are, but not limited to:

> *Relevance of information:* The GDPR principles require that personal data shall be "*adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation')*" (The European Parliament, 2016). Meanwhile, COPPA requires that a privacy notice "*(…) must contain no unrelated, confusing, or contradictory materials*" (Federal Trade Commission, 2013). From our understanding, these requirements relate to Nielsen's 8th usability heuristic "*Aesthetic and minimalist design*", which implies that "*dialogues should not contain information which is irrelevant or rarely needed (…)*" (Nielsen, 2018).
> *Dealing with human error:* GDPR requires that personal data shall be "*accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ('accuracy')*" (The European Parliament, 2016). Similarly, COPPA states that parents "*(…) may refuse to permit the use, and require the deletion, of the information collected*" and the interface must provide a way in which parents can do so (Federal Trade Commission, 2013). As one can compare, these requirements are related to users' control over information availability and accuracy. We understand that these requirements relate to Nielsen's 3rd usability heuristic "*User control and freedom*", which requires that "*Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue (…)*" (Nielsen, 2018).

[1] www.ftc.gov/tips-advice/business-center/guidance/complying-coppa-frequently-asked-questions#General%20Questions.

Usable privacy is a relatively new field, and privacy regulations state premisses that might relate to usability evaluation criteria (e.g., usability heuristics). Usability evaluation criteria support evaluators' judgment to diagnose usability defects[2] on user interfaces. In regards to usable privacy, these defects may implicate on privacy breaches and vulnerabilities (Jøsang et al., 2007; Sasse et al., 2016). The next session describes the Heuristic Evaluation, which is the traditional method to employ usability heuristics as criteria for usability evaluations.

2.2. Heuristic evaluation

Heuristic Evaluation (HE) is a formative usability evaluation method, which aims at diagnosing usability problems at an interface (Lewis, 2014). The method was proposed by Nielsen and Molich (1990). It does not involve the participation of potential users; instead, it involves multiple inspectors that compare the interface against a list of usability heuristics (ISO, 2010).

Although similar to interface guidelines, heuristics are usually small sets of broadly described usability principles, normally around ten heuristics. Therefore, HEs tend to take less time to perform than guidelines review (Lazar et al., 2017). The ten heuristics of Nielsen and Norman are the most traditional set of heuristics employed in HEs. Their titles and descriptions are available at Nielsen Norman Group website (Nielsen, 2018).

Nielsen's revised heuristics (Nielsen, 1994) are traditional in the field. Nevertheless, Nielsen argues that the use of domain specific heuristics may benefit HEs (Nielsen, 1994). Specific domain heuristics have been developed to increase the chances that the usability of software of such domains was not overlooked (Hermawati and Lawson, 2016; de Lima Salgado et al., 2016; de Lima Salgado and Freire, 2014). This is particularly important in the context of Internet of Things (IoT), which raises a variety of new domains to the software industry.

There are different types of usability criteria in the literature. The ISO/IEC 25066 (ISO/IEC, 2016) standard shows that usability criteria include user requirements, guidelines, conventions, style guides, task models and standardized principles. Other terms may be also used as synonym of heuristic. Nevertheless, we agree with Lazar et al. (2017), and understand that a usability criterion may be considered as a heuristic if it is part of a short criteria set (usually no more than ten), and is made to be employed in a HE (shorter in time). In this study, we aimed to find usability heuristics that can be employed in the usability inspection of privacy policy interfaces in the context of smart toys.

2.3. Smart toys, children's privacy protection and privacy policies

2.3.1. Smart toys

Smart toys enhance traditional toys' capabilities with computing services, empowering toys towards ubiquitous computing (de Albuquerque and Kelner, 2019). After an extensive survey of the literature, Albuquerque et al. (2019) shows that there is still no consensus for smart toy terminology. Besides smart toys, they are also called connected toys, interactive toys, toy computing and Internet of Toys (IoToys) (Albuquerque et al., 2019). Despite of all terminologies, these toys can listen and interact with children in new ways. They have recently gained popularity in the market (Mahmoud et al., 2018; Valente and Cardenas, 2017). In this study, we adopt the term smart toy, which is defined by Hung et al. (2017, p. 1) as:

> …a device consisting of a physical toy component that connects to one or more toy computing services to facilitate game play in the Cloud through networking and sensory technologies to enhance the functionality of a traditional toy.

Examples of smart toys are *CogniToys Dino*, *Cue the Robot* and *Sphero BB-8 Robot*; they can talk with children, according to the Mozilla's website "**privacy not included*" (Mozilla, 2018). Another example is *Hello Barbie*, one of the first smart toys in the market. It is a cloud-connected doll that can listen to child's questions and answer them with a cloud-based mechanism. Due to child's privacy concerns, a campaign called "*Hell No Barbie*" was carried out in 2015 by an American advocacy group.[3] To clarify this issue, the Children's Commissioner report (Children's Commissioner, 2018) describes that there are many ways in which child's data, collected by smart toys, can reach wrong hands. As an example, they argued that hackers can gain control over smart toys and talk to children through such devices.

Indeed, as companion robots, smart toys challenge the effectiveness of current privacy regulatory mechanisms (Bertolini and Aiello, 2018) and children privacy protection become a core concept for researches in the field (Chu et al., 2019; Yang et al., 2018; Hung et al., 2018; de Albuquerque and Kelner, 2019; Valente and Cardenas, 2017; Hung et al., 2016; de Albuquerque and Kelner, 2019).

2.3.2. Children's privacy protection

Children's privacy protection refers to enabling "*parents or guardians to be in control of their children's privacy by specifying their privacy preferences for a toy*" (Hung et al., 2018, p. 1). Also, it is crucial to assume that (Rafferty et al., 2017, p. 1227): (i) *children do not understand the concept of privacy and do not know to protect themselves online*, and (ii) *children may disclose personal information to smart toys and not be aware of possible consequences and liabilities*. In this context, parents and guardians strive to protect children's privacy (Rafferty et al., 2017), but there is no universal approach for them to control children's privacy regarding smart toys (Streiff et al., 2019; Rafferty et al., 2017; Xia et al., 2016). Parental privacy controls stand as a promising approach to provide such control and fill the gap, while privacy policies are an essential part of their mechanism (de Lima Salgado et al., 2017; de Azevedo Cunha, 2017; Hung et al., 2018; UNICEF, 2019).

2.3.3. Privacy policies for smart toys

Privacy policies are documents that indicate users' or applications' preferences about data privacy (Garfinkel and Lipford, 2014; Oates et al., 2018). Such preferences indicate the users' choices regarding "*(…) how their PII should be processed for a purpose*" (ISO/IEC, 2011). In other words, privacy policies relate data owners to potential data readers (Jiang and Landay, 2002), and usually control the access of these readers to the data, according to the owner's preferences (Kelley et al., 2009).

In the context of smart toys, parental privacy controls are a promising engine to protect children's privacy (de Lima Salgado et al., 2017; de Azevedo Cunha, 2017; Hung et al., 2018; UNICEF, 2019). Parental privacy controls are features "*for the parents to restrict the content the children can provide to the toy*" (Rafferty et al., 2017, p. 1227). As a type of privacy control, they should have mechanisms for policy generation and interfaces for policy comprehension, policy configuration, and feedback (Paci et al., 2018). Moreover, parental privacy controls must have published accurate privacy policies to ensure children's protection (Hung et al., 2018). For this reason, Yankson et al. (2017)) proposes a privacy preservation framework that supports context-dependent policies based on eXtensible Markup Language (XML). Later on, Yankson et al. (2019) proposes the use of Petri-Nets, a mathematical modeling language, to model, test and verify context-dependent policies.

Privacy policies are usually complicated because they need to represent laws, regulations, and business practices (Schaub et al., 2017). Most of the policies only contain a few options to opt-out and lack details on contextual factors, failing at being effective privacy notice and choice (Schaub et al., 2017; Apthorpe et al., 2019). Also, parents

[2] As per ISO/IEC 25066 (ISO/IEC, 2016)

[3] CBC News at:https://www.cbc.ca/news/business/hello-barbie-1.3292361

may not read these policies nor understand it (Keymolen and der Hof, 2019), and they have no means to verify whether the smart toys follow those policies (Chu et al., 2019). Using *Polisis* (Harkous et al., 2018), a Deep Learning Automated Analysis and Presentation of Privacy Policies, one can see that privacy policies related to smart toy companies are still described with generic security statements, making it difficult for users to understand and chose[4].

Indeed, the usability of privacy policy tools, such as parental privacy controls, challenge the literature (Garfinkel and Lipford, 2014; Paci et al., 2018; Bertino, 2016; De and Zezschwitz, 2016; de Lima Salgado et al., 2017). Although researchers have explored interface alternatives that enhance the usability of privacy policy interfaces (Paci et al., 2018), studies related to parental privacy control for smart toys remain a few. In a previous study of our research group (Harkous et al., 2017), we discussed how traditional HCI methodologies could be applied to enhance the usability of parental privacy controls. In that study, we also proposed initial updates to the design of de Lima Salgado et al.'s (2015) model of parental control aiming to enhance its usability. In a follow-up study (de Lima Salgado et al., 2019), our group proposed a re-design of Rafferty et al.'s parental control model based on Kelley et al.'s (2009) nutrition label model, and a card sort experiment enhanced with cluster analysis. The re-design also contained elements from Google Material Design[5], aiming to enhance usability by interface elements that are familiar to users.

3. Research design

In this paper, we aimed to indicate usability heuristics for the design of parental privacy controls for smart toys. Because parental control for smart toys is a narrow domain, we also evaluated the literature on privacy policy interfaces that are related to smart toys. For this reason, our research design was twofold: we conducted a mapping study (exploratory phase), and an empirical case study (confirmatory phase).

Our mapping study aimed to identify the most appropriate usability heuristics to address usability problems in the domain (broad) of privacy policy interfaces for laypeople. Our empirical case study aimed to confirm the mapping findings in the domain (narrow) of parental privacy controls for smart toys. We performed the case study by empirically comparing the heuristics indicated by the mapping against the traditional usability heuristics of Nielsen. By performing this analysis, we expected to evaluate which set of heuristics was the most appropriated to help professionals in the performance in usability inspections when applied to the specific scenario of parental privacy control.

4. Literature snowballing procedure

This literature mapping performed in the present study followed the snowballing procedure described by Wohlin (Wohlin, 2014). The procedure starts with a brief literature search, which aims at identifying a start set of studies to begin the snowballing. After identifying the start set, the snowballing cycle began. This cycle was divided in two stages: backward and forward. The backward stage evaluates the references indicated at candidate studies, while the forward stage evaluates citations of the candidate studies. These evaluations apply a set of inclusion and exclusion criteria. The snowballing cycle is repeated until saturation, when no new candidate study is identified.

The major advantages of a snowballing procedure is that the backward stage, in most cases, lead straightforward to identify relevant papers. Similarly, using Google Scholar and its citation tracking, the forward stage can be quite informative and helpful to decide about including/excluding a paper (Wohlin, 2014). These characteristics make the snowballing a process which may be focused on the identification of relevant papers rather than measuring the literature. For this reason, the scope of this study was to identify relevant papers that may answer the research question, and then to conduct a follow-up empirical study to analyze the result.

4.1. Defining the start set

We began the mapping with seven (7) candidates for the start set (Jøsang et al., 2007; Jøsang et al., 2007; Nurse et al., 2011; Yeratziotis et al., 2012; Garfinkel and Lipford, 2014; Jaferian et al., 2014; Realpe et al., 2016). Two researchers independently identified these studies because they could potentially answer the research question. To evaluate these candidates and to perform backward and forward stages, we defined the inclusion/exclusion criteria as indicated at Table 1. For the whole procedure, we first compared candidate papers against the exclusion criterion (E1) and, after, against each of the inclusion criteria (I1-I4). Therefore, we only accepted candidate papers that were not excluded by E1 and were included by all inclusion criteria together (I1 AND I2 AND I3 AND I4). We did not define any time frame as inclusion or exclusion criteria; studies from any publication year could be accepted.

We evaluated the seven candidates to the start set against the inclusion/exclusion criteria. Because of criterion I2, we did not accept one candidate (Jøsang et al., 2007), which is referred in this paper as C1. All other candidate papers were accepted and formed the start set as follows:

(S1) A. Jøsang, B. AlFayyadh, T. Grandison, M. AlZomai, J. McNamara, Security Usability Principles for Vulnerability Analysis and Risk Assessment, in: Twenty-Third Annual Computer Security Applications Conference (ACSAC 2007), 2007, pp. 269–278. doi:10:1109/ACSAC:2007:14. URL www:ieeexplore:ieee:org/document/4412995

(S2) J. R. C. Nurse, S. Creese, M. Goldsmith, K. Lamberts, Guidelines for usable cybersecurity: Past and present, in: 2011 Third InternationalWorkshop on Cyberspace Safety and Security (CSS), 2011, pp. 21–26. doi:10:1109/CSS:2011:6058566. URL ieeexplore:ieee:org/document/6058566

(S3) A. Yeratziotis, D. Pottas, D. Van Greunen, A Usable Security Heuristic Evaluation for the Online Health Social Networking Paradigm, International Journal of Human–Computer Interaction 28 (10) (2012) 678–694. doi:10:1080/10447318:2011:654202. URL www:doi:org/10:1080/10447318:2011:654202

(S4) S. Garfinkel, H. R. Lipford, Usable Security: History, Themes, and Challenges, Vol. 5 of SYNTHESIS LECTURES ON INFORMATION SECURITY, PRIVACY, AND TRUST, Morgan & Claypool Publishers, 2014. URL www:ieeexplore:ieee:org/document/6920435

(S5) P. Jaferian, K. Hawkey, A. Sotirakopoulos, M. Velez-Rojas, K. Beznosov, Heuristics for Evaluating IT Security Management Tools, Human–Computer Interaction 29 (4) (2014) 311–350. doi:10:1080/07370024:2013:819198. 240 URL www:dx:doi:org/10:1080/07370024:2013:819198

(S6) P. C. Realpe, C. A. Collazos, J. Hurtado, A. Granollers, A Set of Heuristics for Usable Security and User Authentication, ACM Press, 2016, pp. 1–8. doi:10:1145/2998626:2998662. URL www:dl:acm:org/citation:cfm?doid = 2998626:2998662

From the start set studies (S1–S6), we began the snowballing cycles. We evaluated their references (backward snowballing), and used a Google Scholar mechanism to identify their citations for further evaluation (forward snowballing). We used Google Scholar to avoid publisher bias, as suggested by Wohlin (Wohlin, 2014). We performed these evaluations from August 2nd, 2018 to August 8th, 2018.

[4] We employed pribot.org/polisis tool (security tab) to review ToyTalk.com and zenbo.asus.com privacy policies

[5] material.io/design/

Table 1
Inclusion and exclusion criteria.

Inclusion Criteria
I1. The abstract provides indication of proposal of new usability heuristics for the privacy and security domain.
I2. The full text shows the list of the proposed usability heuristics.
I3. Published in peer-reviewed journals or conferences, or book chapters with editorial boards.
I4. The study **is not** a work in progress or similar unfinished study.
Exclusion Criteria
E1. The study **is not** available in English.

4.2. Iteration 1

At iteration 1, we performed both backward and forward snowballing with the start set. This iteration involved 347 publications (121 from the backward evaluation and 226 from the forward evaluation). Particularly, S2, S3 and S4 are secondary studies, and already synthesized usability heuristics from previous studies. Because our goal was to identify and systematize such heuristics, we did not perform the backward snowballing for S2, S3 and S4.

4.2.1. Backward snowballing

Table 2 indicates the number of candidate studies in this snowballing phase. The first column (*Study*) indicates the source studies for this backward stage. Column *References* shows the number of references retrieved from the respective study/row. Column *Duplicates* indicate whether we identified, and removed, any duplicate from the references. Columns *I1-E1* show the number of references that did not satisfy any of the inclusion/exclusion criteria. Finally, column *New*[6] shows whether any new study satisfied all inclusion/exclusion criteria and composed our mapping.

As indicated at Table 2, this backward stage resulted in the inclusion of one study:

(S7) D. Katsabas, S. Furnell, P. Dowland, Using human computer interaction principles to promote usable security, in: Proceedings of the Fifth International Network Conference (INC 2005), Samos, Greece, 2005, pp. 235–242

The following section presents the outcomes from the forward phase of iteration 1.

4.2.2. Forward snowballing

Table 3 indicates the number of candidate studies for this forward snowballing. It has the same structure as Table 2, despite that it shows a *Citations* column instead of references.

As indicated at Table 3, we included the following studies from this forward snowballing:

(S8) N. Fierro, C. Zapata, Usability Heuristics for Web Banking, in: A. Marcus (Ed.), Design, User Experience, and Usability: Design Thinking and Methods, Springer International Publishing, 2016, pp. 412–423. doi:10:1007/978-3-319-40409-7 39. URL www:link:springer:com/chapter/10:1007/978-3-319-40409-7 39

(S9) C. Altin Gumussoy, Usability guideline for banking software design, Computers in Human Behavior 62 (2016) 277–285. doi:10:1016/j:chb:2016:04:001. URL www:linkinghub:elsevier:com/retrieve/pii/S0747563216302667

(S10) A. Yeratziotis, D. V. Greunen, D. Pottas, Recommendations for usable security in online health social networks, 275 in: 2011 6th International Conference on Pervasive Computing and Applications, 2011, pp. 220–226. doi:10:1109/ICPCA:2011:6106508. URL www:ieeexplore:ieee:org/document/6106508

(S11) G. Reynaga, S. Chiasson, P. C. van Oorschot, Heuristics for the evaluation of captchas on smartphones, ACM Press, 2015, pp. 126–135. doi:10:1145/2783446:2783583. 280 URL www:dl:acm:org/citation:cfm?doid = 2783446:2783583

[6] *New* = *References* − *Duplicates* − *I*1 − *I*2 − *I*3 − *I*4 − *E*1

Table 2
Number of candidate studies involved along *iteration 1 backward snowballing* and new mapped studies.

Study	References	Duplicates	*I1*	*I2*	*I3*	*I4*	*E1*	New
S1	21	1 (C1)	20	0	0	0	0	0
S2[1]	–	–	–	–	–	–	–	0
S31	–	–	–	–	–	–	–	0
S41	–	–	–	–	–	–	–	0
S5	74	0	73	1	0	0	0	0
S6	26	2 (S2 and S5)	22	0	1	0	0	1 (S7)
Totals	121	3 (C1, S2 and S5)	115	1	1	0	0	1 (S7)

[1] S2, S3 and S4 are literature reviews and synthesized usability heuristics from their references. Because our goal was to identify and systematize such heuristics, we did not perform the backward iteration for them.

Table 3
Number of candidate studies involved along *iteration 1 forward snowballing* and new mapped studies.

Study	Citations	Duplicates	*I1*	*I2*	*I3*	*I4*	*E1*	New
S1	69	1 (S2)	61	1	1	0	5	0
S2	49	1 (S6)	46	0	1	0	0	1 (S8)
S3	14	0	10	2	0	0	0	2 (S9 and S10)
S4	44	0	38	2	0	1	3	0
S5	49	4 (S2, S4, S6 and S9)	41	1	0	0	2	1 (S11)
S6	1	0	0	1	0	0	0	0
Totals	226	6	196	7	2	1	10	4 (S8–S11)

At this point, we finalized iteration 1. The following section describes iteration 2, performed with the studies included during this iteration (S7–S11).

4.3. Iteration 2

We performed iteration 2 with the five studies included during iteration 1 (S7–S11). This iteration evaluated 191 candidate papers (145 from the backward evaluation and 46 from the forward evaluation).

4.3.1. Backward snowballing

Table 4 indicates the number of candidate studies along this snowballing phase. Among the five evaluated studies (S7-S11), only S8 returned a new study to the mapping.

This backward snowballing added one study to the mapping, denoted as:

(S12) F. Paz, F. A. Paz, J. A. Pow-Sang, L. Collantes, Usability Heuristics for Transactional Web Sites, in: 2014 11th International Conference on Information Technology: New Generations, 2014, pp. 627–628. doi:10:1109/ITNG:2014:81. URL www:ieeexplore:ieee:org/document/6822272

The following section presents the forward snowballing of iteration 2.

4.3.2. Forward snowballing

Table 5 indicates the number of studies involved along this snowballing phase. Similarly to iteration 2 backward, only S8 returned a new study to the mapping.

This forward snowballing added one study to the mapping, denoted

Table 4
Number of candidate studies involved along *iteration 2 backward snowballing* and new mapped studies.

Study	References	Duplicates	*I1*	*I2*	*I3*	*I4*	*E1*	New
S7	14	0	13	0	1	0	0	0
S8	20	1 (S2)	17	0	1	0	0	1 (S12)
S9	37	1 (S3)	35	0	1	0	0	0
S10	17	0	17	0	0	0	0	0
S11	57	0	55	0	2	0	0	0
Totals	145	2 (S2 and S3)	137	0	5	0	0	1 (S12)

Table 5
Number of candidate studies involved along *iteration 2 forward snowballing* and new mapped studies.

Study	Citations	Duplicates	*I1*	*I2*	*I3*	*I4*	*E1*	New
S7	15	1 (S6)	14	0	0	0	0	0
S8	2	0	0	0	0	0	1	1 (S13)
S9	14	0	11	0	0	0	3	0
S10	6	0	6	0	0	0	0	0
S11	9	0	7	1	0	0	1	0
Totals	46	1 (S6)	0	0	0	0	0	1 (S13)

as:

(S13) G. Baños Díaz, C. M. d. P. Zapata Del Río, A Proposal of Usability Heuristics Oriented to E-Banking Websites, in: A. Marcus, W. Wang (Eds.), Design, User Experience, and Usability: Theory and Practice, Vol. 10918, Springer International Publishing, Cham, 2018, pp. 327–345. doi:10:1007/978–3-319–91797-9 23. URL www:link:springer:com/10:1007/978–3-319–91797-9_23

The next section presents the third iteration, performed after.

The next section presents the third iteration, performed after the outcomes of this one.

4.4. Iteration 3

We performed iteration 3 with the two new studies included after iteration 2 (S12 and S13). This iteration evaluated 61 candidate papers (42 from backward evaluation and 19 from forward evaluation).

4.4.1. Backward snowballing

Table 6 shows the number of studies involved at this backward snowballing. As indicated, this backward snowballing did not include any new study to the mapping.

4.4.2. Forward snowballing

Table 7 indicates the number of studies involved at iteration 3 forward snowballing. As shown, no study was included after this forward snowballing. Because both backward and forward stages of iteration 3 did not include any new study to this mapping, the snowballing procedure was finished. The following section presents the data extraction and the analysis.

At this phase, we concluded the collection of new studies for our mapping. The following section present and discuss the mapping results.

Table 6
Number of candidate studies involved along *iteration 3 backward snowballing* and new mapped studies.

Study	References	Duplicates	***I1***	***I2***	***I3***	***I4***	***E1***	New
S12	8	0	8	0	0	0	0	0
S13	34	2 (S8 and S12)	32	0	0	0	0	0
Totals	42	2 (S8 and S12)	40	0	0	0	0	0

Table 7
Number of candidate studies involved along *iteration 3 forward snowballing* and new mapped studies.

Study	Citations	Duplicates	***I1***	***I2***	***I3***	***I4***	***E1***	New
S12	19	2 (S8 and S13)	13	0	0	0	4	0
S13	0	–	–	–	–	–	–	0
Totals	19	2 (S8 and S13)	0	0	0	0	0	0

5. Results from the mapping study

To indicate usability heuristics for the design of parental privacy controls for smart toys, we mapped 13 studies from the literature (S1–S13). Each of these studies suggests usability criteria to employ in HEs of security and privacy systems. Our results include studies from 2005 up to 2018. Because we did not limit a time frame for this mapping, these results suggest that the first study to propose usability principles for the security and privacy domain dates back to 2005. Therefore, it indicates that this research topic may have 14 years, which would represent an average of almost one mapped study per year. Nevertheless, eight out of the 14 studies dates from 2014 to 2019, which may indicate an increased interest on the topic. Fig. 1 indicates the mapped studies along the timeline of this research topic.

We extracted data from the 13 studies and evaluated against the themes proposed by Hermawati and Lawson (Hermawati and Lawson, 2016, p. 35) to assess domain specific usability heuristics. The themes (T#) are: "*Adequacy of the domain*" (T1), "*Creation process*" (T2), "*Validation*" (T3), "*Adequacy of heuristics' description*" (T4) and "*Effectiveness*" (T5). These five themes guided our discussion of the results, and the evaluation of the research question.

5.1. Adequacy of the domain

Among the 13 included studies, we identified heuristics for eight different domains: *Usable privacy and security*, *Banking software*, *Access control*, *Information Technology Security Management (ITSM)*, *User authentication*, *Online Health Social Network (OHSN)*, *captchas on smartphones(Completely Automated Public Turing test to tell Computers and Humans Apart* [7]*)* and *transactional websites*. Table 8 relates these domains with their respective studies. As indicated, none of the studies proposed heuristics for lay privacy policy and smart toys.

In a broad sense, we found generic usability principles for the cyber-security domain (S1, S2, S3 and S7) and for cyber-security subdomains (S4, S5, S6, S8, S9, S10, S11, S12 and S13). Among the usability principles for cyber-security subdomains, S4 and S5 have closest focus on the domain of privacy controls. S4 proposes heuristics for access controls, which require users to view and author privacy policies and is close related to our domain of interest (Garfinkel and Lipford, 2014). Meanwhile, S5 proposes heuristics for ITSM. ITSM involve different user profiles, such as software developers, security auditors and lay users. Therefore, these heuristics may also be appropriate for the domain of lay privacy policy interfaces and smart toys. With regards to this theme, it may indicate that S4 and S5 sets are more appropriate to evaluate usability of privacy policy interfaces.

5.2. Heuristics definition process

Among the 13 studies (S1–S13), we identified ten different definition methodologies for the sets of heuristics. The method proposed by Rusu et al. (2011) is predominant among these studies (S6, S8, S12 and S13). This method comprises six steps to establish usability heuristics:

[7] *Definition retrieved from: dictionary.cambridge.org/dictionary/english/captcha*

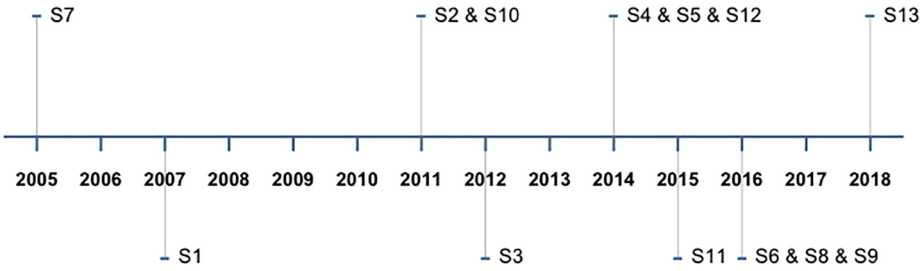

Fig. 1. Timeline of usability principles for security and privacy tools.

Table 8
Domains identified from mapped studies.

Domain	Study
Usable privacy and security	S1, S2, S3 and S7
Banking software	S8, S9 and S13
Access control	S4
Information Technology Security Management (ITSM)	S5
User authentication	S6
Online Health Social Network (OHSN)	S10
Captchas	S11
transactional websites	S12

exploratory, descriptive, correlational, explicative, validation and refinement. It is important to notice that S6 only adopted the first four (4) of these steps (Realpe et al., 2016), neither validation nor refinement steps were performed.

Meanwhile, the method proposed by Yeratziotis et al. (2011) is the second most referred. Both **S3** and **S10** adopted this methodology. The method of Yeratziotis et al. is a three-phase process. Phase 1 is to design high-level heuristics, phase 2 is the validation of the high-level heuristics and phase 3 the application of these high-level heuristics. These are not cascade phases, and researchers can go back and forth among them. It is just not possible to go straight from phase 1 (design) to phase 3 (application). This method is not only similar to Rusu et al.'s (2011), but also more recent. Such similarity indicates that there is a need for a standardized methodology to create usability heuristics in the field. Yet, it reinforces the recommendations of Hermawati and Lawson (2016) that the creation of usability heuristics should contemplate validation and refinement stages.

Six studies defined their own method to create their heuristics. Three of them (S5, S7 and S9) conducted some type of validation during the process. In S5, Jaferian et al. defined their own method based on grounded theory techniques (Jaferian et al., 2014). They performed a top-down methodology to justify, support and combine design guidance into new usability heuristics. They also performed empirical validation and refinement stages.

At S7, Katsabas et al. (2005) adapted Johnston et al.'s (2003) usability criteria together with Nielsen's heuristics to compose their own. During this process, they also considered aspects of the cyber-security domain and the first principles of interaction design (Tognazzini, 2014). This process was, in some extent, subjective and centered on the author's knowledge about the cyber-security domain.

In study S9, Altin Gumussoy (Altin Gumussoy, 2016) adapted the heuristics of Muller et al. (1998). She evaluated the coverage of these heuristics against a database of three banking software projects. She also performed a cluster analysis to group heuristics according to its coverage of usability problems by severity.

The remaining three studies (S1, S2 and S4) defined their own method to create the heuristics, without validation stages. S1 reports limited information about how their principles were created (Jøsang et al., 2007), but informed that such principles were created from the Kerckhoffs' principles for identity management (Kerckhoffs, 1883). S2 reviewed the literature on usability recommendations for cyber-security systems and consolidated a set of 19 guidelines (Nurse et al., 2011). Nurse et al. (2011) grouped similar recommendations into unique guidelines, and renamed such guidelines in accordance to the recommendation content. S4 also summarized previous literature to compose their set of guidelines (Garfinkel and Lipford, 2014). Their process is described as a subjective analysis of the authors about the literature reviewed. After, they summarized lessons learned from the literature as guidelines to create usable access control mechanisms.

To sum up, nine out of the 13 studies (S3, S5, S7, S8, S9, S10, S11, S12 and S13) performed some type of validation of their principles during its creation. As recommended by Hermawati and Lawson (2016), and supported by our results, including validation procedures during the creation of usability principles may enhance its quality. For this reason, these nine studies stand as more appropriate than the others to be employed in further HEs in the privacy and security domain. The following section present further description of validation processes among the mapped studies.

5.3. Validation

Nine of the studies reported some type of validation (S3, S5 and S7–S13) when proposing their new heuristics. Five of them (S5, S8, S11, S12 and S13) empirically compared their heuristics against Nielsen's. The other four (S3, S7, S9 and S10) performed alternative validation methods. We understand that those validation procedures that empirically compared new heuristics against Nielsen's are more appropriate to our propose. The following sections describe each of these methods.

5.3.1. Comparison with Nielsen's heuristics

Studies S5, S8, S11, S12 and S13 compared their heuristics with Nielsen's heuristics. At **S5**, Jaferian et al. (2014) designed a between-subjects study with 28 participants (inspectors), equally divided between two groups. Both groups evaluated the same interface, from one identity management system. Most of the participants performed remote evaluation. All of them had, at least, HCI and computer security background, and previous experience with HE.

In **S8**, Fierro and Zapata (2016) describe a between group comparison. They do not inform the number of inspectors in each group, neither inspector background. In their case study, the subject for evaluation was one international bank website.

In **S11**, Reynaga et al. (2015) conducted a between group study with 18 participants. They were divided in two groups of nine, each group performing HEs with different heuristics. One group employed Nielsen's heuristics, while the other employed the new heuristics for captchas on mobile phones. The evaluators were HCI experts, also familiar with computer security. Reynaga et al. (2015) chose four captchas images to be evaluated during the HEs. These captchas were chosen because they represent the main captchas categories.

At **S12**, Paz et al. (2014) organized four HEs with different groups of inspectors. The study does not indicate the number of inspectors in each group. Nevertheless, three groups employed Nielsen's heuristics, while the other one employed the new heuristics. They evaluated one exemplar transactional website. The study does not provide information about inspectors' background.

S13 reports a validation process comparing their new heuristics against Nielsen's. At such study, Díaz and Zapata (Baños Díaz et al., 2018) designed a between group comparison; one group applied Nielsen's heuristics while the other applied the new ones. A baking website was subject of inspection. The study does not inform the number of inspectors in each group, neither the inspectors' background.

5.3.2. Alternative validation methods

This section overview studies that did not compared their heuristics against Nielsen's (S3, S7, S9 and S10). Among these studies, S3 demonstrated the application of their heuristics, while the others (S7, S9 and S10) performed some kind of questionnaire-based validation. None of them described conducting a HE with the proposed heuristics.

At **S3**, Yeratziotis et al. (2012) demonstrated the application of their heuristics against a set of examples. They did not count on any external evaluators to perform the validation. Also, they used one interface of Online Health Social Networking for the demonstrations.

During ***S7***, Katsabas et al. (2005) applied their principles in a questionnaire-based evaluation (not a HE). Inspectors were required to related each principle to a five-point scale ("*Application diverges completely from the guideline*" to "*Application completely follows the guideline in all possible sections*"). No description of inspectors' background was given. Ten interfaces were subject of evaluation. According to the authors, these interfaces could give an overall mix of both security-specific tools.

In **S9**, Altin Gumussoy (2016) counted on three usability experts to rate the severity of 266 known usability problems. These problems came from three banking software projects. After, the same experts indicated how well each of the new principles describe each of the problems. To perform this stage, the experts should answer a six-point scale questionnaire. This research design aimed to reveal the interaction between severity level of usability problems and each of the new principles. Also, it allowed the author to cluster their principles according to the severity of problems that each of the principles is related to.

Finally, at **S10**, Yeratziotis et al. (2011) conducted a study with six evaluators. All evaluators were postgraduate students in the field of Information and Communication Technology. The evaluators were requested to inspect two online health social networks. They were provided with scenarios for the evaluators. After, the evaluators applied the new heuristics and rated a five-point scale questionnaire ("*Very Good*" to "*Very Poor*") to indicate the usability of the interface.

Table 9
Studies, description style and number of principles (N).

Study	Description	***N***
S1	Succinct	8
S2	Traditional	19
S3	Lengthy	86
S4	Succinct	5
S5	Traditional	7
S6	Succinct	75
S7	Traditional	10
S8	Traditional	15
S9	Lengthy	51
S10	Traditional	7
S11	Traditional	7
S12	Traditional	15
S13	Traditional	6

5.4. Adequacy of heuristics' description

This mapping identified 278 usability heuristics. Seven studies proposed sets with no more than ten heuristics (S4 = 5; S13 = 6; S5 = 7; S10 = 7; S11 = 7; S1 = 8; S7 = 10). Meanwhile, six studies proposed sets with more than ten rules (S3 = 13; S9 = 13; S8 = 15; S12 = 15; S2 = 19; S6 = 153). Although S6 presents a set of 153 principles, only 75 are considered by the authors as related to usability (Realpe et al., 2016). Despite of the number of heuristics, studies also varied on how to describe them. Among the studies, we identified the three description styles:

Succinct: the description is composed by a title (or succinct description. This is the case of checkpoints and quick tips.
Traditional: the description is composed by a title and one or two paragraphs. We called this as traditional because it is the closest to the description of Nielsen's traditional heuristics.
Lengthy: the description is composed by a title and multiple checklist items (Yeratziotis et al., 2012) or multiple usability criteria (Altin Gumussoy, 2016). This style increases the total number of rules to be considered during a HE, which may be not desired.

Table 9 indicates each study and relates them to their respective description style and the number (N) of criteria. We identified three sets with succinct descriptions (S1, S4 and S6), eight with traditional descriptions (S2, S5, S7, S8, S10, S11, S12 and S13) and two sets with lengthy descriptions (S3 and S9). Because S3 and S9 have lengthy descriptions, they have an increased number of criteria. This would make the number of criteria in S3 increases from 13 to 86, and from 13 to 51 in S9. We understand that such large number is due to the lack of standardization on heuristic description styles. However, large number of heuristics may be inappropriate to be employed in HEs, which is often intended to be fast, compared to guideline reviews (Lazar et al., 2017).

As indicated by Lazar et al. (2017), we understand that heuristic sets with no more than ten (10) heuristics are more appropriate to be employed in HEs, especially considering that inspectors are meant to memorize them and to apply them in the process, and not have to constantly go back to their descriptions as one would do in a review of guidelines. With this characteristic, seven studies stand as more appropriate heuristic descriptions (S1, S4, S5, S7, S10, S11 and S13).

5.5. Effectiveness of the heuristics

Five studies (S5, S8, S11, S12 and S13) measured and discussed the effectiveness of their new criteria. Some studies adopted the traditional effectiveness metric, as showed by Hartson et al. (2001), while others adopted alternative measures. S5 adopted the traditional effectiveness metric, and S11 provided enough information to calculate it as well. **S5** found that their heuristics had higher *f-measure* compared with Nielsen's heuristics (S5 = 0.80; Nielsen's = 0.72; α = 0.5) (Jaferian et al., 2014). **S11** argues that their heuristics resulted in more unique usability problems and less false positives than Reynaga et al.'s (2015). Based on the values they reported, we calculated the *f-measure* of their heuristics. We found that it had higher *f-measure* than Nielsen's heuristics (S11 = 0.82; Nielsen's = 0.73; α = 0.5). These results are close to the results shown by S5.

On the other hand, S8, S12 and S13 adopted alternative measures. **S8** argues that their new criteria resulted on the uncovering of 46% of unique problems, while Nielsen's resulted in 34% (Fierro and Zapata, 2016). **S12** presents a case study where their new heuristics found more usability problems than Nielsen's. However, they could not make any inference about the observed advantage. **S13** shows that Nielsen's heuristics were more efficient (discovery of usability problems by time) them their new heuristics (Baños Díaz et al., 2018). Although, they argued that their heuristics resulted in a higher number of usability problems (56%) than Nielsen's (40%). Also, they showed that their heuristics resulted in identification of more severe problems.

In summary, the mapped studies indicated that Nielsen's heuristics have an effectiveness about 0.72 and 0.73 for HEs in the privacy and security domain. Meanwhile, they indicated that domain-focused heuristics had effectiveness rates of about 0.80 and 0.83 for HEs in the privacy and security domain.

6. Comparison between domain-specific and general usability heuristics on the inspection of parental privacy control of smart toys

To confirm the findings from the mapping study with preliminary empirical evidence, we conducted a case study with 20 participants[8]. We sought to compare the diagnosis of usability problems between Nielsen's heuristics and the ITSM heuristics.

To this end, we invited 20 novice usability inspectors based on feasibility analysis (Caine, 2016). They were all undergraduate students in Computer Science that, at the time this study was conducted, had just completed their Human–Computer Interaction course. We decided to invite such novice inspectors because experts are rare to find and have limited available time to participate as voluntary in our study. Studying the performance of novice inspectors may help the field to enhance their performance in HEs, which may benefit the organizations that relies on their performance. Therefore, the quasi-experimental design (between subjects) is defined as presented at Table 10.

Based on the parental control model proposed by de Lima Salgado et al. (2015), we designed a prototype be analyzed at the inspections. To populate the prototype with real world information, we used the information from an online Brazilian toy store.[9] We made the prototype in Portuguese, because the participants were all native speakers of Portuguese (Brazilians). All participants had one hour to complete the inspection and deliver their individual list of problems. They were not required to perform group discussion after the HE. To ensure that participants would find a minimum number of usability problems to diagnose, we introduced known usability problems (incorrect words and disabled fields) to the prototype ("*seeding known usability problems*" (Hartson et al., 2001, p. 384)).

Table 10
Planned quasi-experimental design (between subjects).

Nielsen's condition	ITSM condition
10 participants	10 participants

Table 11
Feasible quasi-experimental design (between subjects).

Condition	Participants (p#)	Diagnosed Problems (N)
Nielsen	p1	5
Nielsen	p2	12
Nielsen	p3	7
Nielsen	p4	7
Nielsen	p5	7
Nielsen	p6	6
Jaferian et al.	p7	11
Jaferian et al.	p8	13
Jaferian et al.	p9	8
Jaferian et al.	p10	8
Jaferian et al.	p11	16
Jaferian et al.	p12	13
Jaferian et al.	p13	9
Jaferian et al.	p14	8

Although the 20 participants voluntarily accepted to participate in our study and began the inspection, six of them decided to quit their participation before completing the inspection. For this reason, we did not include their inspections' results. This fact left us with the feasible quasi-experimental design as presented at Table 11.

Overall, the six participants that used Nielsen's heuristics diagnosed 24 usability problems ($Min.$=5;$Median$ = 7;$Mean$ = 9.667;$sd \approx$ 5.0;$Max.$=17). After a Shapiro–Wilk normality test ($p - value \approx 0.038$), we found that these results may have been drawn from a normal distribution. Meanwhile, the eight participants that used the ITSM heuristics diagnosed 29 usability problems ($Min.$=8;$Median$ = 11.50;$Mean$ = 13.12; $sd \approx$ 4.36;$Max.$=19); and a Shapiro–Wilk normality test ($p - value \approx 0.13$) has shown that we cannot assume, based on the observations of this study, that the ITSM would be normally distributed.

To compare both conditions, we produced a standard usability problem set by the union of usability problem sets obtained from the empirical HEs ("*Union of usability problem sets over UEMs*[10] *being compared*" (Hartson et al., 2001, p. 384)). To this end, we applied a Kruskal–Wallis rank sum test to compare both distributions according to the number of diagnosed usability problems. The results show that the ITSM heuristics had a significantly higher coverage of problems in the diagnosis of usability problems (Kruskal–Wallis chi-squared = 6.1381, df = 1, p-value $\approx$ 0.01), which is in accordance with the findings of our mapping study. Nevertheless, future studies may compare these heuristic sets using outcomes from test with users as a benchmark set. Test with users may also reveal usability problems that were not diagnosed by the inspectors, and indicate if there is a lack of coverage of problems among the heuristics. In case a lack of coverage is verified, new heuristics might be necessary to cover them.

To calculate the f-measure of each condition, we used the formulas as indicated by Hartson et al. (2001, p. 390-394):

[8] This study was approved by a Research Ethics Committee with CAAE code 69353317.4.0000.5390

[9] We have the permission of *Clube Reborn* & *Toys* to use their content for research purposes.

[10] *UEMs: Usability Evaluation Methods*

$$Thoroughness = \frac{number\ of\ diagnosed\ problems}{size\ of\ the\ benchmark\ set} \quad (1)$$

$$F - measure = \frac{1}{\alpha\left(\frac{1}{Validity}\right) + (1-\alpha)\left(\frac{1}{Thoroughness}\right)} \quad (2)$$

In sequence, we aggregated the usability problems diagnosed by both groups (37 problems) to become the benchmark set. Because we adopted the union of both conditions as the benchmark set, the validity measure for both conditions was equal to one (1), making the f-measure equal to the thoroughness. The results also show that the Jaferian et al.'s condition influenced a higher f-measure (≈ 0.78) compared to Nielsen's condition (≈ 0.65). Therefore, these results also confirms the findings of our map, evidencing that the ITSM heuristics are the most appropriate for lay privacy policy interfaces in the domain of smart toys.

After the HEs, we asked the inspectors about their opinion on the method. All inspectors affirmed that they had difficulties to assign the heuristic to problems during the inspection, which is similar to the feedback from inspectors in the study of Jaferian et al. (2014), when inspectors let to assign the heuristics after identifying the problems. For this reason, and because we limited the inspection time on one hour, none of the participants assigned heuristics to their diagnosed problems.

Although inspectors that used the ITSM heuristics diagnosed a higher number of usability problems, there were eight (8) usability problems identified only by inspectors that used Nielsen's heuristics. These problems refer to general usability aspects, which are not specific related to smart toys' privacy. Meanwhile, among the 13 problems that were only diagnosed by those inspectors using the ITSM heuristics, there are both general usability problems and privacy related problems. Moreover, inspectors that employed the ITSM heuristics found all the eight (8) usability problems related to smart toys' privacy, while those that employed Nielsen's heuristics found only three of them. The full list of usability problems, with the indication of which group diagnosed them, is presented at Appendix A.

Table 12 presents the usability problems related to smart toys' privacy. For each of them, we analyzed which of the ITSM heuristics is most appropriate to address the respective problem. Interestingly, the problem "*The application does not follow conventions for user authentication*" was diagnosed by inspectors that employed the ITSM heuristics. However, the ITSM heuristics do not cover consistency related problems. This fact emphasizes the need for a consistency heuristic among the ITSM heuristics, which was also observed during their study ("*The need for a consistency heuristic was indicated by PI2, PI14, and PI6.*" (Jaferian et al., 2014, p. 341).

Table 12
Coverage of ITSM heuristics on observed usability problems that are related to smart toys' privacy

Jaferian et al.'s Heuristic	Usability Problems
Rules and constraints	- At the "*Obligations and Retention*" screen, it is only possible to select PIPEDA and stated purpose - At the "*Obligations and Retention*" screen, it is only possible to select PIPEDA and stated purpose - At the "*Review and add rule*" screen, there is no option to disagree with the policy - At the "*Child Information*"screen, users can only go ahead if they agree with the terms
Visibility of activity status	- At the "*Review and add rule*", the privacy rule description is still complex to understand - At the "*Review Privacy Policy*", at the Clube Reborn policies, the unique contact information is a WhatsApp number (external app).
Planning and dividing work between users	- The application does not provide feedback on what user is using the application (father/mother/guardian)
None	- The application does not follow conventions for user authentication

Table 13
Summary of adequacies between studies and themes (T#).

Study	T1	T2	T3	T4	T5
S1				x	
S2					
S3		×			
S4	×			×	
S5	×	×	×	×	×
S6					
S7		×		×	
S8		×	×		
S9		×			
S10		×		×	
S11		×	×	×	×
S12		×	×		
S13		×	×	×	

Finally, as indicated at Table 12, only three out of the seven ITSM heuristics were necessary to refer to usability problems related to smart toys' privacy, while one of these problems could not be referred by their heuristics. This may indicate the need for creating new heuristics to cover all usability problems related to smart toys' privacy (e.g. privacy consistency and standards). Future studies may explore this gap.

7. Conclusions

This study aimed to answer the question: *What usability heuristics best address usability problems that affect laypeople interaction with privacy policy interfaces related to smart toys?* To answer this question, we performed a snowballing mapping study, evaluating 589 publications (despite potential duplicates) among three snowballing iterations (337 + 191 + 61), which resulted in 13 mapped studies (S1–S13).

Table 13 summarizes the results of our map and their adequacy with the evaluated themes (T#). We indicate studies that are adequate to the particular theme with a "×". Among the mapped studies, no one proposed heuristics for our domain (T1) of interest (privacy policy interfaces for smart toys). Yet, two studies (S4 and S5) suggested heuristics for broad domains that include lay privacy policy interfaces for generic security devices. Nine studies[11] describe the creation of their heuristics (T2) with some validation process (T3). Five of them[12] empirically compared their heuristics against Nielsen's, which may better support its employment on HEs. Finally, two studies (S5 and S11) reported traditional metrics that indicate the effectiveness of their proposed heuristics.

As shown in Table 13, the heuristics indicated in **S5** stand as the most appropriate to answer our question. To confirm this finding, we conducted a case study comparing the heuristics proposed in S5 (ITSM heuristics) against the traditional heuristics of Nielsen. The results of the case study confirmed that the ITSM heuristics have a greater and significant impact on the diagnosis of usability problems in lay privacy controls for smart toys (Kruskal–Wallis, p-value ≈ 0.01). For this reason, we conclude that, to the extent to which this study covered, the ITSM heuristics (Jaferian et al., 2014) best-addressed problems that affect laypeople's interaction with parental privacy controls for smart toys, and can be pointed as the answer to our research question. Nevertheless, future studies may explore the creation of new heuristics for lay privacy policy controls for smart toys, and compare against

[11] S3, S5, S7, S8, S9, S10, S11, S12, and S13
[12] S5, S8, S11, S12, and S13

Nielsen's and the ITSM heuristics.

Based on the findings of this study, we suggest as future studies to compare Nielsen's and the ITSM heuristic sets using outcomes from tests with users as a source of a benchmark set. We also recommend exploring the creation of new domain-specific heuristics for parental privacy controls for smart toys, which may include aspects of consistency that are not described among the ITSM heuristics.

CRediT authorship contribution statement

André de Lima Salgado: Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Project administration, Funding acquisition. **Renata Pontin de Mattos Fortes:** Conceptualization, Methodology, Resources, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition. **Ricardo Ramos de Oliveira:** Methodology, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing. **André Pimenta Freire:** Methodology, Resources, Writing - original draft, Writing - review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was supported by the grants 2015/24525-0, 2017/15239-0 and 2018/26038-8, São Paulo Research Foundation (FAPESP). This study was funded in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil* (CAPES) – Finance Code 001. This study was supported by The Brazilian National Council for Scientific and Technological Development (CNPq-MCTIC).

Appendix A. List of usability problems

Table A.14
List of usability problems.

Usability Problems	ITSM	Nielsen
Many text input fields are disable along the screens	×	×
The scroll is not working on the "*Review Privacy Policy*" screen	×	×
It is not possible to check the absolute location and, then, the relative location on the "*Core Access Control*" screen.	×	×
At the "*Core Access Control*", it is not possible to click next after selecting only the relative location	×	×
At the "*Core Access Control*", all checkbox are unchecked if users uncheck the "*Read*" checkbox	×	
Many checkbox are disabled among the interface screens.	×	×
At the "*Parent/Guardian Detials*", it is not possible to go forward after selecting "*No, please do not send updates*"	×	×
At the "*Parent/Guardian Detials*", users need to uncheck a checkbox before checking the other one	×	×
Among all the screens, the back button undo users' action instead of going back to the previous screen	×	
The first screen should not have a return (back) button		×
At the "*Purposes*" screen, only one combination of checkbox (game and administrative) allows users to go forward with the task.	×	×
At the "*Recipients*" screen, it is not possible to select "*Anyone*"	×	×
At the "*Recipients*" screen, when users uncheck an option, all the other options are unchecked together	×	×
At the "*Obligations and Retention*" screen, it is only possible to select PIPEDA and stated purpose	×	
The interface does not provide descriptions for the abbreviations used	×	×
Navigation with tabs is not working	×	×
At the "*Obligations and Retentions*" screen, the search field is disabled	×	
The screen title and the blue menu present the same information (repetitive).		×
On the Obligations and Retention screen, there is no efficient way to select multiple options.		×
The application does not follow conventions for user authentication	×	
At the create new rule screen, the application does not prevent users from advancing without inserting necessary information (e.g. unable the next button)		×
At the "*create a new rule*" pop up, the model option is not working	×	
At the "*create a new rule*"screen, before the pop up, the "next" button is unnecessary	×	×
At the "*Review and add rule*" screen, there is no option to disagree with the policy	×	×
The application does not provide appropriate error messages to users.		×
When users finish their task, the application does not return to the home screen.		×
"*Legal Guardian*" is still a technical term for laypeople.		×
At the first screen, the word child is incorrectly written.		×
The interface does not provide descriptions for the jargon	×	×
There is no indication (feedback) when the task is ended	×	
At the "*Manage privacy rules*" screen, users can only check the absolute location checkbox	×	
At the "*Review and Finish*" screen, it is only possible to uncheck the email option	×	
The application does not provide feedback on what user is using the application (father/mother/guardian)	×	
At the "*Review and add rule*", the privacy rule description is still complex to understand	×	×
At the "*Child Information*" screen, users can only go ahead if they agree with the terms	×	
At the "*Review Privacy Policy*", at the *Clube Reborn* & *Toys* policies, the unique contact information is a WhatsApp number (external app)	×	
At every screen, it is still difficult to understand whether a button is enabled or disabled	×	

References

Acquisti, A., Taylor, C.R., Wagman, L., 2016. The Economics of Privacy, SSRN Scholarly Paper, Social Science Research Network, Rochester, NY. doi:10.2139/ssrn.2580411. URL: www.papers.ssrn.com/abstract=2580411.

Albuquerque, D.O.P., Fantinato, M., Kelner, J., Albuquerque, A.P.d., 2019. Privacy in smart toys: Risks and proposed solutions. Electronic Commerce Research and Applications, 100922. https://doi.org/10.1016/j.elerap.2019.100922. URL: www.sciencedirect.com/science/article/pii/S1567422319300997.

Alonso-Ríos, D., Mosqueira-Rey, E., Moret-Bonillo, V., 2018. A systematic and generalizable approach to the heuristic evaluation of user interfaces. Int. J. Human–Comput. Int. 1–14. https://doi.org/10.1080/10447318.2018.1424101.

Altin Gumussoy, C., 2016. Usability guideline for banking software design. Comput.

Human Behav. 62, 277–285. https://doi.org/10.1016/j.chb.2016.04.001. URL: www.linkinghub.elsevier.com/retrieve/pii/S0747563216302667.
Apthorpe, N., Varghese, S., Feamster, N. 2019. Evaluating the contextual integrity of privacy regulation: Parents' iot toy privacy norms versus coppa, 123–140. doi:10.5555/3361338.3361348.
Baños Díaz, G., Zapata Del Río, C.M. d. P., 2018. A Proposal of usability heuristics oriented to e-banking websites. In: In: Marcus, A., Wang, W. (Eds.), Design, User Experience, and Usability: Theory and Practice, vol. 10918. Springer International Publishing, Cham, pp. 327–345 10.1007/978-3-319-91797-9_23.
Bertino, E., 2016. Data security and privacy: concepts, approaches, and research directions. In: 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC), vol. 1. pp. 400–407 URL: www.ieeexplore.ieee.org/document/7552042.
Bertolini, A., Aiello, G., 2018. Robot companions: a legal and ethical analysis. Inf. Soc. 34 (3), 130–140. https://doi.org/10.1080/01972243.2018.1444249.
Caine, K., 2016. Local Standards for Sample Size at CHI. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems+, CHI '16. ACM, New York, NY, USA, pp. 981–992 doi: 10.1145/2858036.2858498.
Children's Commissioner, 2018. Who knows what about me? A Children's Commissioner report into the collection and sharing of children's data, Tech. rep. URL: www.childrenscommissioner.gov.uk/our-work/digital/who-knows-what-about-me/.
Chu, G., Apthorpe, N., Feamster, N., 2019. Security and privacy analyses of internet of things children's toys. IEEE Internet Things J. 6 (1), 978–985. https://doi.org/10.1109/JIOT.2018.2866423. URL www.ieeexplore.ieee.org/document/8443103.
Cranor, L.F., Buchler, N., 2014. Better together: usability and security go hand in hand. IEEE Security Privacy 12 (6), 89–93. https://doi.org/10.1109/MSP.2014.109. URL: www.ieeexplore.ieee.org/document/7006405.
de Albuquerque, A.P., Kelner, J., 2019. Toy user interfaces: systematic and industrial mapping. J. Syst. Architect. https://doi.org/10.1016/j.sysarc.2018.12.001. URL: www.sciencedirect.com/science/article/pii/S138376211830153X.
de Albuquerque, A.P., Kelner, J., 2019. Non-personal data collection for toy user interfaces. In: Proceedings of the 52nd Hawaii International Conference on System Sciences, https://doi.org/10.24251/HICSS.2019.209. URL: www.scholarspace.manoa.hawaii.edu/handle/10125/59612.
De, L.A., Zezschwitz, E.V. 2016. Usable privacy and security, it - Information Technology 58 (5) 215–216. doi:10.1515/itit-2016-0034. URL www.degruyter.com/view/j/itit.2016.58.issue-5/itit-2016-0034/itit-2016-0034.xml.
de Azevedo Cunha, M.V., 2017. Unicef, others, Child Privacy in the Age of Web 2.0 and 3.0: Challenges and opportunities for policy, Tech. Rep. URL www.unicef-irc.org/publications/926-child-privacy-in-the-age-of-web-20-and-30-challenges-and-opportunities-for-policy.html.
de Lima Salgado, A., Freire, A.P., 2014. Heuristic evaluation of mobile usability: A mapping study. In: M. Kurosu (Ed.), Human-Computer Interaction. Applications and Services, Springer International Publishing, Cham, pp. 178–188. doi:10.1007/978-3-319-07227-2_18. URL www.link.springer.com/chapter/10.1007/978-3-319-07227-2_18.
Rafferty, L., Fantinato, M., Hung, P.C.K., 2015. Privacy requirements in toy computing. In: Hung, P.C.K. (Ed.), Mobile Services for Toy Computing. Springer International Publishing, pp. 141–173 URL: www.link.springer.com/chapter/10.1007/978-3-319-21323-1_8.
de Lima Salgado, A., Rodrigues, S.S., Fortes, R.P.M., 2016. Evolving heuristic evaluation for multiple contexts and audiences: perspectives from a mapping study. In: Proceedings of the 34th ACM International Conference on the Design of Communication, SIGDOC '16. Association for Computing Machinery, New York, NY, USA 10.1145/2987592.2987617.
de Lima Salgado, A., do Amaral, L.A., Castro, P.C., de Mattos Fortes, R.P., 2017. Designing for Parental Control: Enriching Usability and Accessibility in the Context of Smart Toys. Springer International Publishing, Cham, pp. 103–125 Doi: 10.1007/978-3-319-62072-5_7.
de Lima Salgado, A., do Amaral, L.A., Castro, P.C., de Mattos Fortes, R.P., 2017. Designing for parental control: enriching usability and accessibility in the context of smart toys. In: Tang, J.K., Hung, P.C.K. (Eds.), Computing in Smart Toys. Springer International Publishing, Cham, pp. 103–125 doi:10.1007/978-3-319-62072-5_7.
de Lima Salgado, A., Dias, F.S., Mattos, J.a.P.R., de Mattos Fortes, R.P., Hung, P.C.K., 2019. Smart toys and children's privacy: usable privacy policy insights from a card sorting experiment. In: Proceedings of the 37th ACM International Conference on the Design of Communication, SIGDOC '19. Association for Computing Machinery, New York, NY, USA 10.1145/3328020.3353951.
EU, 2018. Handbook on European data protection law, 2018th Edition, Handbook/ FRA, European Union (EU) Agency for Fundamental Rights, European Court of Human Rights, Council of Europe and European Data Protection Supervisor, Publications Office of the European Union, Luxembourg, 2018.
Federal Trade Commission ("FTC" or "Commission"), 2013. Children's online privacy protection rule, The Office of the Federal Register (OFR) of the National Archives and Records Administration (NARA), and the U.S. Government Publishing Office (GPO) at the FederalRegister.gov website.
Fierro, N., Zapata, C., 2016. Usability Heuristics for Web Banking. In: In: Marcus, A. (Ed.), Design, User Experience, and Usability: Design Thinking and Methods Springer International Publishing, pp. 412–423 doi: 10.1007/978-3-319-40409-7_39.
Garfinkel, S., Lipford, H.R., 2014. Usable security: history, themes, and challenges. Synthesis Lectures on Information Security, Privacy, and Trust, vol. 5 Morgan & Claypool Publishers URL: www.ieeexplore.ieee.org/document/6920435.
Harkous, H., Fawaz, K., Lebret, R., Schaub, F., Shin, K.G., Aberer, K., 2018. Polisis: Automated analysis and presentation of privacy policies using deep learning. In: Proceedings of the 27th USENIX Conference on Security Symposium, SEC'18. USENIX Association, USA, pp. 531–548 10.5555/3277203.3277243.
Hartson, H.R., Andre, T.S., Williges, R.C., 2001. Criteria for evaluating usability evaluation methods. Int. J. Human–Comput.Interact. 13 (4), 373–410. https://doi.org/10.1207/S15327590IJHC1304_03.
Hermawati, S., Lawson, G., 2016. Establishing usability heuristics for heuristics evaluation in a specific domain: is there a consensus? Appl. Ergonom. 56, 34–51. https://doi.org/10.1016/j.apergo.2015.11.016. URL: www.sciencedirect.com/science/article/pii/S0003687015301162.
Hung, P.C., Iqbal, F., Huang, S.-C., 2016. Children's privacy protection engine for smart anthropomorphic toys. Eng. Med. Appl. 15.
Hung, P.C.K., Tang, J.K.T., Kanev, K., 2017. Introduction. In: J.K. Tang, P.C.K. Hung (Eds.), Computing in Smart Toys, Springer International Publishing, Cham, 2017, pp. 1–5. doi:10.1007/978-3-319-62072-5_1. URL www.doi.org/10.1007/978-3-319-62072-5_1.
Hung, P.C.K., Fantinato, M., Roa, J., 2018. Children Privacy Protection. In: Lee, N. (Ed.), Encyclopedia of Computer Graphics and Games, Springer International Publishing, Cham, pp. 1–3. doi:10.1007/978-3-319-08234-9_198-1. URL www.doi.org/10.1007/978-3-319-08234-9_198-1.
ISO, 2010. ISO 9241-210:2010, Ergonomics of human-system interaction – part 210: Human-centred design for interactive systems. International Organization for Standardization (ISO), URL: www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-1:v1:en.
ISO/IEC, 2011. ISO/IEC 29100:2011, Information technology – Security techniques – Privacy framework. International Organization for Standardization (ISO), URL: www.iso.org/obp/ui/#iso:std:iso-iec:29100:ed-1:v1:en.
ISO/IEC, 2016. ISO/IEC 25066:2016, Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Common Industry Format (CIF) for Usability – Evaluation Report, Tech. rep. International Organization for Standardization (ISO), URL: www.iso.org/obp/ui/#iso:std:iso-iec:25066:ed-1:v1:en.
ISO, 2018. ISO 8124-1:2018(en), Safety of toys – Part 1: Safety aspects related to mechanical and physical properties. International Organization for Standardization (ISO), URL: www.iso.org/obp/ui/#iso:std:iso:8124:-1:ed-5:v1:en.
ISO/TR, 2018. ISO/TR 18638:2018, health informatics – guidance on health information privacy education in healthcare organization. International Organization for Standardization (ISO), URL: www.iso.org/obp/ui/#iso:std:iso:tr:18638:ed-1:v1:en.
Jaferian, P., Hawkey, K., Sotirakopoulos, A., Velez-Rojas, M., Beznosov, K., 2014. Heuristics for evaluating IT security management tools. Human–Comput. Interact. 29 (4), 311–350. https://doi.org/10.1080/07370024.2013.819198.
Jang-Jaccard, J., Nepal, S., 2014. A survey of emerging threats in cybersecurity. J. Comput. Syst. Sci. 80 (5), 973–993. https://doi.org/10.1016/j.jcss.2014.02.005. URL: www.sciencedirect.com/science/article/pii/S0022000014000178.
Jiang, X., Landay, J.A., 2002. Modeling privacy control in context-aware systems. IEEE Pervasive Comput. 1 (3), 59–63. https://doi.org/10.1109/MPRV.2002.1037723.
Johnston, J., Eloff, J., Labuschagne, L., 2003. Security and human computer interfaces. Comput. Security 22 (8), 675–684. https://doi.org/10.1016/S0167-4048(03)00006-3.
Jøsang, A., AlFayyadh, B., Grandison, T., AlZomai, M., McNamara, J., 2007. Security usability principles for vulnerability analysis and risk assessment. In: Twenty-Third Annual Computer Security Applications Conference (ACSAC 2007), pp. 269–278 URL www.ieeexplore.ieee.org/document/4412995.
Jøsang, A., Zomai, M.A., Suriadi, S., 2007. Usability and privacy in identity management architectures. In: Proceedings of the Fifth Australasian Symposium on ACSW Frontiers - Volume 68, ACSW '07. Australian Computer Society Inc, 152, pp. 143 doi: 10.5555/1274531.1274548.
Juniper Research, 2017. Smart Toy Sales to Grow Threefold to Exceed $15.5 Billion by 2022, Tech. rep. URL: www.juniperresearch.com/press/press-releases/smart-toy-sales-to-grow-threefold.
Katsabas, D., Furnell, S., Dowland, P., 2005. Using human computer interaction principles to promote usable security. In: Proceedings of the Fifth International Network Conference (INC 2005). Samos, Greece, pp. 235–242.
Kaushik, K., Kumar Jain, N., Kumar Singh, A., 2018. Antecedents and outcomes of information privacy concerns: role of subjective norm and social presence. Electron. Commer. Res. Appl. 32, 57–68. https://doi.org/10.1016/j.elerap.2018.11.003.
Kelley, P.G., Bresee, J., Cranor, L.F., Reeder, R.W., 2009. A nutrition label for privacy. In: Proceedings of the 5th Symposium on Usable Privacy and Security, SOUPS '09. ACM, New York, NY, USA, pp. 4:1–4:12 doi: 10.1145/1572532.1572538.
Kerckhoffs, A., 1883. La cryptographie militaire IX 5–38.
Keymolen, E., der Hof, S.V., 2019. Can i still trust you, my dear doll? A philosophical and legal exploration of smart toys and trust. J. Cyber Policy 4 (2), 143–159.
Lazar, J., Feng, J.H., Hochheiser, H., 2017. Research Methods in Human–Computer Interaction. Morgan Kaufmann, Cambridge, MA, USA.
Lewis, J.R., 2014. Usability: lessons learned...and yet to be learned. Int. J. Human–Comput. Interact. 30 (9), 663–684. https://doi.org/10.1080/10447318.2014.930311.
Mahmoud, M., Hossen, M.Z., Barakat, H., Mannan, M., Youssef, A., 2018. Towards a comprehensive analytical framework for smart toy privacy practices. In: Proceedings of the 7th Workshop on Socio-Technical Aspects in Security and Trust, STAST '17. Association for Computing Machinery, New York, NY, USA, pp. 64–75.
McReynolds, E., Hubbard, S., Lau, T., Saraf, A., Cakmak, M., Roesner, F., 2017. Toys that listen: a study of parents, children, and internet-connected toys. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 5197–5207 doi: 10.1145/3025453.3025735.
Mozilla, 2018. Privacy Not Included: A Buyer's Guide for Connected Products. URL: www.foundation.mozilla.org/en/privacynotincluded/categories/Toys & Games/.
Muller, M.J., Matheson, L., Page, C., Gallup, R., 1998. Methods & tools: participatory heuristic evaluation. Interactions 5 (5), 13–18. https://doi.org/10.1145/285213.285219.
Nielsen, J., 1994. Enhancing the explanatory power of usability heuristics. In:

Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '94. ACM, New York, NY, USA, pp. 152–158 10.1145/191666.191729.
Nielsen, J., 2018. Heuristic evaluation: How to conduct a heuristic evaluation. URL: www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/.
Nielsen, J., Molich, R., 1990. Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, pp. 249–256 10.1145/97243.97281.
Nurse, J.R.C., Creese, S., Goldsmith, M., Lamberts, K., 2011. Guidelines for usable cybersecurity: past and present. 2011 Third International Workshop on Cyberspace Safety and Security (CSS) 21–26. https://doi.org/10.1109/CSS.2011.6058566.
Oates, M., Ahmadullah, Y., Marsh, A., Swoopes, C., Zhang, S., Balebako, R., Cranor, L.F., 2018. Turtles, locks, and bathrooms: understanding mental models of privacy through illustration. Proc. Privacy Enhancing Technol. 4https://doi.org/10.1515/popets-2018-0029. URL: www.content.sciendo.com/view/journals/popets/2018/4/article-p5.xml.
Paci, F., Squicciarini, A., Zannone, N., 2018. Survey on access control for community-centered collaborative systems. ACM Comput. Surv. 51 (1), 6:1–6:38. https://doi.org/10.1145/3146025.
Paz, F., Paz, F.A., Pow-Sang, J.A., Collantes, L., 2014. Usability Heuristics for Transactional Web Sites. In: 2014 11th International Conference on Information Technology: New Generations, pp. 627–628. https://doi.org/10.1109/ITNG.2014.81. URL: www.ieeexplore.ieee.org/document/6822272.
Rafferty, L., Hung, P., Fantinato, M., Marques Peres, S., Iqbal, F., Kuo, S.-Y., Huang, S.-C., 2017. Towards a privacy rule conceptual model for smart toys. In: Proceedings of the 50th Hawaii International Conference on System Sciences, https://doi.org/10.24251/HICSS.2017.146. URL: www.scholarspace.manoa.hawaii.edu/handle/10125/41299.
Realpe, P.C., Collazos, C.A., Hurtado, J., Granollers, A., 2016. A Set of Heuristics for Usable Security and User Authentication. ACM Press 1–8. https://doi.org/10.1145/2998626.2998662.
Reynaga, G., Chiasson, S., van Oorschot, P.C., 2015. Heuristics for the Evaluation of Captchas on Smartphones. ACM Press 126–135. https://doi.org/10.1145/2783446.2783583. URL: www.dl.acm.org/citation.cfm?doid=2783446.2783583.
Rusu, C., Roncagliolo, S., Rusu, V., Collazos, C., 2011. A methodology to establish usability heuristics. In: Proceedings of The Fourth International Conference on Advances in Computer-Human Interactions (ACHI), pp. 4 URL: www.hcibib.org/ACHI11.
Sasse, M.A., Smith, M., 2016. The security-usability tradeoff myth [guest editors' introduction]. IEEE Security Privacy 14 (5), 11–13. https://doi.org/10.1109/MSP.2016.102.
Sasse, M.A., Smith, M., Herley, C., Lipford, H., Vaniea, K., 2016. Debunking security-usability tradeoff myths. IEEE Security Privacy 14 (5), 33–39. https://doi.org/10.1109/MSP.2016.110. URL: www.ieeexplore.ieee.org/document/7676175.
Schaub, F., Balebako, R., Cranor, L.F., 2017. Designing effective privacy notices and controls. IEEE Internet Comput. 21 (3), 70–77. https://doi.org/10.1109/MIC.2017.75. URL: ieeexplore.ieee.org/document/7927931.
Still, J.D., 2016. Cybersecurity needs you!. Interactions 23 (3), 54–58. https://doi.org/10.1145/2899383.
Streiff, J., Das, S., Cannon, J., 2019. Overpowered and underprotected toys empowering parents with tools to protect their children. In: IEEE Humans and Cyber Security Workshop (HACS 2019).
The European Parliament and The Council of The European Union, Regulation (EU) 2016/679 of the European parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (2016).
Tognazzini, B., 2014. First Principles of Interaction Design (Revised & Expanded). URL www.asktog.com/atc/principles-of-interaction-design/.
UNICEF, 2019, Memorandum on Artificial Intelligence and Child Rights. UNICEF Innovation, Human Rights Center, UC Berkeley, URL www.unicef.org/innovation/reports/memoAIchildrights.
Valente, J., Cardenas, A.A., 2017. Security & privacy in smart toys. In: Proceedings of the 2017 Workshop on Internet of Things Security and Privacy, IoTS&P '17. ACM, New York, NY, USA, pp. 19–24. https://doi.org/10.1145/3139937.3139947.
Valente, J., Cardenas, A.A., 2017. Security & privacy in smart toys. In: Proceedings of the 2017 Workshop on Internet of Things Security and Privacy, IoTS&P '17. Association for Computing Machinery, New York, NY, USA, pp. 19–24 10.1145/3139937.3139947.
Wash, R., Zurko, M.E., 2017. Usable security. IEEE Internet Comput. 21 (3), 19–21. https://doi.org/10.1109/MIC.2017.69. URL: www.ieeexplore.ieee.org/document/7927871.
Wohlin, C., 2014. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, EASE '14. ACM, New York, NY, USA, pp. 38:1–38:10 doi: 10.1145/2601248.2601268.
Xia, Z., Wang, X., Sun, X., Wang, Q., 2016. A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data. IEEE Trans. Parallel Distrib. Syst. 27 (2), 340–352. https://doi.org/10.1109/TPDS.2015.2401003.
Yang, J., Lu, Z., Wu, J., 2018. Smart-toy-edge-computing-oriented data exchange based on blockchain. J. Syst. Architect. 87, 36–48. https://doi.org/10.1016/j.sysarc.2018.05.001. URL: www.sciencedirect.com/science/article/pii/S1383762118300638.
Yankson, B., Iqbal, F., Hung, P.C.K., 2017. Privacy Preservation Framework for Smart Connected Toys. Springer International Publishing, Cham pp. 149–164.
Yankson, B., Iqbal, F., Lu, Z., Wang, X., Hung, P., 2019. Modeling privacy preservation in smart connected toys by petri-nets. In: Proceedings of the 52nd Hawaii International Conference on System Sciences, https://doi.org/10.24251/HICSS.2019.207.
Yeratziotis, A., Greunen, D.V., Pottas, D., 2011. Recommendations for usable security in online health social networks. In: 2011 6th International Conference on Pervasive Computing and Applications, pp. 220–226. https://doi.org/10.1109/ICPCA.2011.6106508. URL: www.ieeexplore.ieee.org/document/6106508.
Yeratziotis, A., Pottas, D., van Greunen, D., 2011. A three-phase process to develop heuristics. In: Proceedings of the 13th ZA-WWW Conference.
Yeratziotis, A., Pottas, D., Van Greunen, D., 2012. A usable security heuristic evaluation for the online health social networking paradigm. Int. J. Human–Comput. Interact. 28 (10), 678–694. https://doi.org/10.1080/10447318.2011.654202.