Abstract
The development of usable interfaces for privacy policies is essential to increase users’ trust in technology and comply with legal requirements. This thesis aimed to design interfaces that allow laypeople to protect their online privacy. A comprehensive analysis was conducted, comprising a literature review, a thematic and cluster analysis, and an empirical evaluation. Six usable privacy heuristics (push) were derived, which effectively detect severe problems in privacy policy interfaces for laypeople. Moreover, initial usable privacy guidelines (pug) were formulated, and a novel process for developing usability criteria was proposed. Future research directions were suggested, such as applying these heuristics and guidelines to domains like human-robot interaction and human-artificial intelligence interaction.

Keywords: usable privacy, heuristic, heuristic evaluation, usability, inspection, security

1 Introduction
Governments worldwide have introduced privacy legislation, such as the European Union General Data Protection Act (GDPR), the California Consumer Privacy Act (CCPA), the Canadian Personal Information Protection and Electronic Documents Act (PIPEDA), the Japanese Act on Protection of Personal Information (APPI), and the Brazilian General Personal Data Protection Law (LGPD) [32]. These regulations emphasize the need for individuals and organizations to obtain appropriate authorization before handling personal data. In the context of information technology, this implies that users must be informed and given the freedom to decide whether to share their personal information with other entities [32].

To inform users about data collection and usage, information technology companies commonly present privacy policies within their software applications [1, 32]. These policies often include privacy choices and settings that allow users to determine how their personal data is shared.

primary threat to information security [34]. However, as the spectrum of potential threats expanded due to the ubiquity of data [2], everyday individuals were increasingly tasked with making security decisions [22], and they emerged as a pivotal factor in the field’s advancement [34]. It became evident that even experts could inadvertently misconfigure systems and leave vulnerabilities without user-friendly tools [31]. Consequently, blame for security breaches might shift from users to designers [37]. Despite this evolution, certain fundamental aspects of how laypeople interact with privacy tools, such as mental models [25], remain underexplored in the literature. The UPS field continues to require user-friendly tools tailored for laypeople [2], along with suitable usability methodologies catered to this specific domain.

Examples of prominent themes within the UPS domain encompass, but are not confined to, social media privacy, user authentication, anti-phishing endeavors, Web privacy, and equitable information practices. This research endeavor centers on Web privacy and equitable information practices, concentrating on the challenges associated with crafting privacy policy tools that are user-friendly for individuals without specialized knowledge. The impetus behind addressing Web privacy and equitable information practices arises from the escalating opportunities for data collection facilitated by the Internet. Online retailers, in particular, gained unprecedented capabilities to amass and scrutinize data pertaining to their clientele [15]. Consequently, privacy regulations stipulate that users must possess awareness and agency in deciding whether to share their personal information with external entities [32]. To this end, the U.S. Government mandated companies to safeguard users’ privacy through disclosure and choice mechanisms [15]. Similarly, the General Data Protection Regulation (GDPR) compelled enterprises within the European Union to empower individuals with control over their data sharing activities on the Web [35]. Notably, the Brazilian Government also upholds Web users’ privacy rights 1. Subsequent sections delve into the usability aspects of privacy policy tools.

Parallel to usability, information security is deeply intertwined with industry benchmarks. While defining usability remains a challenge in scholarly discourse, defining information security frequently revolves around upholding the confidentiality, availability, and integrity of information. As technology advances, new attributes may augment these core aspects to better address the intricacies of information security processes [36]. Consequently, it is reasonable to conceive of the term “usable security” as the endeavor to render the information security process user-friendly. However, it is important to acknowledge that this perspective doesn’t encompass the entirety of the usable privacy and security domain.

If usable security is to design usable security processes, we should assume that users’ goal is to secure the system. But, “security is usually a secondary goal” for laypeople [38]. The unmotivated user property, as defined by Whitten and Tygar [38], states that:

Security is usually a secondary goal. People do not generally sit down at their computers wanting to manage their security; rather, they want to send an email, browse web pages, or download software, and they want security in place to protect them while they do those things. It is easy for people to put off learning about security, or to optimistically assume that their security is working, while they focus on their primary goals. Designers of user interfaces for security should not assume that users will be motivated to read manuals or to go looking for security controls that are designed to be unobtrusive. Furthermore, if security is too difficult or annoying, users may give up on it altogether.

This thesis delves into an exploration of the usability of privacy policy interfaces tailored for individuals without technical expertise, specifically targeting the safeguarding of children’s privacy, termed as parental privacy control. The underlying premise is that the desire to manage the privacy of their loved ones serves as a motivational factor encouraging laypeople to engage with privacy policy interfaces.

Commonly, privacy policy tools encompass a combination of interfaces and mechanisms for policy creation, comprehension, configuration (referred to as privacy choice), and feedback provision. These mechanisms encapsulate the primary components of privacy policy tools [26]. As outlined by Paci et al. [26], a significant portion of research pertaining to policy creation centers around automated or semi-automated policy generation methods. Studies concerning policy comprehension frequently introduce novel interface designs, a strategy mirrored in endeavors aimed at enhancing policy comprehension interfaces. Conversely, investigations into feedback generation primarily revolve around furnishing feedback pertinent to access decision-making [26]. This work concentrates on research related to privacy policy comprehension and configuration/choice, given that within the literature, these studies predominantly introduce fresh interface designs for this domain.

3 Methods and Results

To develop the usability heuristics, we performed a thematic analysis of 45 transcripts retrieved from the literature, associating them with higher-level themes [3, 5]. These transcripts consist of user feedback and provide empirical indications of user behavior with privacy policy tools. Two researchers independently coded the initial themes by identifying user

---

behaviors in the transcripts, continuing the process until theoretical saturation was reached. A total of 19 initial themes were identified, representing the main usability issues encountered by laypeople when interacting with privacy policy tools. These themes were then revised to suggest usable privacy recommendations for future designs of privacy policy tools. The resulting recommendations serve as design requirements for the development of privacy policy tools and can be seen as a preliminary set of "Privacy and Usability Guidelines (pug’s)".

After achieving saturation of the initial themes, we revisited the 45 transcripts to record the occurrences of the 19 initial themes. A cluster analysis was then conducted on this dataset to identify higher-level themes that could form the basis of the new heuristics. The 19 privacy and usability guidelines (pug’s) were used as objects in the cluster analysis, with their occurrences serving as attributes. Ward’s method and Jaccard’s distance were applied to cluster the objects, aiming to minimize within-group dispersion and avoid weighting negative matches. A maximum of ten clusters was sought to correspond to a maximum of ten heuristics, ensuring a clear distinction from usability guidelines. Ultimately, six clusters were identified, forming the basis of the developed usable privacy heuristics, referred to as “Privacy and Usability Heuristics (push)”. These heuristics were numbered from 1 to 6 based on their explanatory power within the data. Heuristics with higher explanatory power were listed first. The six usable privacy heuristics, which we call Privacy and Usability Heuristics (push#) are:

- **push#1 Readability of privacy policies.**
  The readability of privacy policies is crucial for users to understand how they share their data. Users may want to access personalized privacy analysis to understand the risks of sharing their data. While users set their privacy choices, they become vigilant and start to explore the interface and assess policies. They may also prefer interfaces with fewer privacy choices instead of complex settings.

- **push#2 Users’ doubt and precaution.**
  Users may assess the consequences of others’ privacy choices before deciding about their own.

- **push#3 Provide help and avoid jargon.**
  Users may search for specific information, such as terms and definitions, or seek help to understand privacy policies. Avoid jargon.

- **push#4 Discretionary access control.**
  Users may want to know the extent to which their personal data is being shared. They may also want to know who accesses their personal data. After that, they may want to restrict access to their data.

- **push#5 Fast interaction and human error vulnerabilities.**
  Users may seek fast interactions with privacy policies. To this end, they deduce how the interface works. In these cases, they need to quickly understand how the privacy choice settings work (conceptual model), and human error is very likely to occur.

- **push#6 Unstable choices and appropriate symbols.**
  Users may change their privacy choices over time. A good policy-choice mapping is desirable in these situations. Employing appropriate symbols enhances the mapping.

To evaluate the effectiveness of the push# heuristics for privacy controls, a quasi-experimental study with a between-group design was conducted, comparing these new heuristics against the state-of-the-art usability heuristics proposed by Jaferian et al. [8, 21]. Start-up professionals were invited to participate voluntarily in the study, and they were randomly assigned to either the control group (using Jaferian et al.’s heuristics) or the treatment group (using the push# heuristics). The sample size for each group was determined following the recommendations of Caine [4], aiming for statistical power with Cohen’s $d = 0.5$, $\alpha = 0.05$, and $\beta = 0.85$. Based on these requirements, a minimum of four participants in each group was required. The evaluation focused on the parental privacy control model proposed by Rafferty et al. [27], which was prototyped and accessed online via a browser simulation of a mobile device. The performance of the heuristics was compared based on the downstream utility, a measure of their ability to generate valuable outputs for the design change process. Usability problems identified by participants were categorized according to their severity using a scale proposed by Yankson et al. [39], with a focus on privacy-related issues.

**Downstream utility on catastrophic problems**

We compared the downstream utility on the number of catastrophic problems diagnosed by participants in each group. In the control group, participants discovered a minimum of zero and a maximum of two catastrophic usability problems ($\bar{X} = 0.375, M = 0.00, S \approx 0.42$). Meanwhile, in the treatment group, participants discovered a minimum of zero and a maximum of three catastrophic usability problems ($\bar{X} = 0.9375, M = 1.00, S \approx 1.00$). After a normality test, both the results’ distribution from the control group $(p-value = 9.986 - 0.06)$ and from the treatment group $(p-value \approx 0.0065)$ significantly differed from a normal distribution. Therefore, as indicated by Lazar et al. [23], we employed the Wilcoxon signed-rank test to statistically compare the downstream utility of the number of catastrophic problems between groups. The results are shown in Figure 1 with jitter, boxplots, and the respective statistical results. For the downstream utility on catastrophic problems, evaluators’ performance using the push# heuristics was significantly greater than the performance of evaluators using the state-of-the-art ($p-value = 0.038$). In addition, the data analysis showed a moderate effect size ($0.317$).

For instance, the results of push# heuristics’ performance on catastrophic problems evidence its benefits over the state-of-the-art. For this reason, we accept the thesis hypothesis for the scope of catastrophic problems.

**Downstream utility on major problems**

Comparing the downstream utility on the number of major problems reported, participants in the control group discovered a minimum of zero and a maximum of four major


Figure 1. Results of the Wilcoxon signed-rank test.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Total of diagnosed catastrophic problems</th>
<th>Wilcoxon test, ( W = 86, p = 0.038, n = 32 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group</td>
<td>( \bar{X} = 1.00, M = 0.00, S \approx 1.37 )</td>
<td>( p - value \approx 0.0008 )</td>
</tr>
<tr>
<td>Treatment group</td>
<td>( \bar{X} = 1.562, M = 1.50, S \approx 1.82 )</td>
<td>( p - value \approx 0.004 )</td>
</tr>
</tbody>
</table>

In the control group, participants discovered a minimum of zero and a maximum of six major problems (\( \bar{X} = 1.562, M = 1.50, S \approx 1.82 \)). After a normality test, both results’ distribution from the control group significantly differed from a normal distribution. Therefore, we compared the number of diagnosed problems between groups by applying the Wilcoxon signed-rank test. Although the observed results from the treatment group suggest a slight improvement compared with the control group, both groups had similar performance in this measure.

### Overall number of usability problems reported

In addition to comparing the downstream utility of the heuristics, we compared the number of usability problems diagnosed by participants in each group. In the control group, participants discovered a minimum of zero and a maximum of five usability problems (\( \bar{X} = 1.75, M = 1.00, S \approx 1.95 \)). Meanwhile, in the treatment group, participants discovered a minimum of zero and a maximum of eight usability problems (\( \bar{X} = 3.125, M = 3.00, S \approx 2.66 \)). After a normality test, the results’ distribution from the control group significantly differed from a normal distribution (\( p - value \approx 0.002 \)), while the results’ distribution from the treatment group did not (\( p - value \approx 0.14 \)). Therefore, we compared the number of diagnosed problems between groups by applying the Wilcoxon signed-rank test. Although the results from the treatment group suggest an overall improvement compared with the control group, the difference was not significant.

### Number of minor and cosmetic problems reported

Comparing the number of minor problems reported, participants in the control group discovered a minimum of zero and a maximum of one minor problem (\( \bar{X} = 0.375, M = 0.00, S = 0.5 \)). Meanwhile, in the treatment group, participants discovered a minimum of zero and a maximum of two minor problems (\( \bar{X} = 0.375, M = 0.00, S \approx 0.62 \)). After a normality test, both results’ distribution from the control (\( p - value = 2.566e - 05 \)) and treatment (\( p - value = 4.803e - 05 \)) group significantly differed from a normal distribution. Therefore, we compared the number of diagnosed problems between groups by applying the Wilcoxon signed-rank test. Although the treatment group’s observed results suggest a slight improvement compared with the control group, both groups had similar performance in this measure.

Similar to the number of minor problems, we observed a slight difference in the number of cosmetic problems reported between groups. Participants in the control group did not find any cosmetic problems. Meanwhile, in the treatment group, participants discovered a minimum of zero and a maximum of three cosmetic problems (\( \bar{X} = 0.25, M = 0.00, S \approx 0.77 \)). After a normality test, the results’ distribution from the treatment group did not (\( p - value = 2.444e - 07 \)).

The data analysis on the discovery of cosmetic problems suggests that the push# heuristics have higher coverage over the ITSM. However, the observed results are not significant to confirm it.

### 4 Conclusions and Contributions

This doctoral research began with exploratory studies in game and web accessibility, usability inspection for novice evaluators, and methods for comparing usability finding reports [9, 13, 16, 29]. These studies emphasized the importance of supporting software development teams with context information to improve user interface evaluation outcomes and highlighted the challenges of comparing usability findings.

Thereafter, state-of-the-art usability heuristics were identified for evaluating privacy policy interfaces [8]. Pilot studies and experiments were conducted to create usable privacy recommendations, assess user behavior in the privacy context, and pilot methodology [10, 11, 14, 30]. The results provided valuable insights and contributed to the creation of usability criteria and guidelines.

In the second iteration of the PhD planning stage, users’ behavior with privacy protection interfaces was further explored, including studies on smart toys, Data Glove interfaces, and connected-autonomous vehicles [10, 14, 30]. Our usability criteria creation method was piloted using grounded theory techniques [11]. Also, we revisited the severity rating criteria for usability problems found on privacy policy control applications [39].

In the final executing stage, new usable privacy criteria were created and evaluated, focusing on parental privacy control interfaces of smart toys. Overall, this doctoral research delivered nine papers and one book section to the literature as is. Besides, it also generated one registered software.

The comprehensive and extended version of this paper can be found in the referenced study by Salgado et al. [28]. This extensive rendition delves deeper into the research presented in the initial paper, providing a more intricate analysis, additional experimental results, and an enriched discussion of...
the findings. Readers interested in gaining a more comprehensive understanding of the subject matter are encouraged to refer to this extended version, which offers a more comprehensive and detailed exploration of the concepts and insights originally introduced.

**Acknowledgments**

This study was supported by the grants 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.

**References**


