H. Roßnagel et. al. (Eds.): Open Identity Summit 2019,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 185

Lessons learned – Conducting a User Experience evaluation of a Trust Policy Authoring Tool

Stephanie Weinhardt[1], Doreen St. Pierre[2]

Abstract: Most contributions on usable policy authoring and usable IT-Security only focus on the design phase of a tool and on stating guidelines how to make these tools and systems user friendly. There are only some contributions introducing work regarding usability evaluations but even less introducing user experience evaluations. This contribution wants to address this lack. Based on a user experience evaluation with a trust policy authoring tool we present the lessons learned derived from the results.

Keywords: user experience evaluation, trust policy authoring, evaluation methods, lessons learned.

1 Introduction

Although a lot of work on usability in IT-Security has been conducted [ZSS96], [CG04], [Bo06], [MJ08], [FRR09], [KS14], [Pr17], [Ia18] as mostly expert users or administrators deal with creating privacy or security policies, user experience (ux) seems to be no requirement. Therefor, the necessety to develop user friendly tools remains a secondary goal. But with recent developments in IoT, automation, industry 4.0 and the overall increasing connectivity between humans and machines, the need for not only usable but also positively experienced policy authoring tools increases [ZSS96], [Bo06], [CI06], [FIM10]. Users need to be enabled to formulate their own privacy, security and trust policies to ensure the protection of their data and interests.

An essential part in ensuring high usability is the evaluation of a concept [Hu10]. But existing contributions mostly consider the design phase, creating a lack of contributions on the evaluation phase and used methods and approaches [Vo17].

To the best of our knowledge, this contribution conducted the first ux evaluation of a (trust) policy authoring tool. The evaluation was conducted with a high-fidelity prototype, which was developed within the European funded research project LIGHT^{est}. One of the project goals is to enable all kinds of users to create their own trust policies. To meet this reguirement a three-layered approach based on [KBK05], [Br06], [CI06] was implemented. Each layer provides a different way to create a policy, tailored to the user groups: novice, intermediate and expert users. To meet the needs of each, each layer has

[1] Universität Stuttgart IAT, Institut für Arbeitswissenschaft und Technologiemanagement, Competence Team Identity Management, Allmandring 35, 70569 Stuttgart, stephanie.weinhardt@iao.fraunhofer.de
[2] Fraunhofer IAO, Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO, Competence Team Identity Management, Allmandring 35, 70569 Stuttgart, doreen.stpierre@iao.fraunhofer.de

a different set of functionalities presented in a tailored way. This results in not overstraining novice users that want to create simple and basic policies, but still provide full functionality to expert users. We included ux evaluation methods, to investigate if the users have a positive experience with the tool and what basic human needs are causing it.

2 Evaluation setting

In the prototype only the layer for novice users and for intermediate users was implemented. The approach was to have a standard laboratory usability test with a preliminary questionnaire, task completion with the prototype and a follow-up questionnaire. To include ux measurements we added the User Needs Questionnaire (UNeeQ) [FP14], as well as follow-up interviews focusing individually on the ux of each participant. The UNeeQ measures to which extent a set of ten basic human needs are addressed and how the overall user experience with a product was. It measures that by letting participants rate their accordance to predefined sentences on a five level likert-scale from "not at all" to "highly". As these measurements are highly subjective and depend on the current mood of the participants, they had to fill out the UNeeQ right before and directly after interacting with the prototype. Thus, the data could be compared as we measured the users' mindset before the interaction and how the interaction changed it. In combination with the follow-up interviews we hoped for insights to which needs are addressed.

The evaluation was conducted with 18 participants with an average duration of 60-75 minutes. All participants were able to complete the tasks with no major usability problems. Despite the overall positive usability, the UNeeQ showed contradictory results. Comparing the results from the UNeeQ before the interaction and the results after the interaction, every value for each need decreased between 0,15 and 1,25 (on a scale from 0 to 4). We explain that reduced experience with the limited autonomy that could be experienced with the prototype and missing to create a more realistic and elaborate scenario.

Although we mixed usability and ux evaluation methods and there are short comings in the evaluation setting, we nevertheless, wanted to reflect on the lessons learned.

3 Lessons learned

We integrated ux measurements into the usability evaluation to gather first experiences in the area of IT-Security and policy authoring. In the following we state the lessons learned.

"The maturity level of the prototype, the context of use and the evaluation setting are highly dependent on each other for a successful ux evaluation."

We figured that having a high-fidelity prototype would be sufficient for an ux evaluation

in which we wanted to learn about the basic human needs being addressed while interacting with it. But the fact that participants could not explore the system freely on their own and had to execute predefined tasks caused hindered participants autonomy and therefore hindered an experience to develop. Thus, we would recommend a highly developed prototype or functioning system for an ux laboratory evaluation. This was also described by [La17]. If you do not have a highly developed prototype/ system available, you could show participants a video with the intended usage scenario and ask them to think of the experience they anticipate with the system. Of course this is a subjective, reflected and anticipated opinion on participants' side, but delivers more insight into participants thinking and which basic human needs could be addresses.

"Include valence method combined with laddering interviews."

Although the follow-up interview included questions about the good and the bad aspects of the tool, as well as what the users perceived as easy and not easy we got little to know about the experience and the basic human needs. As the users had already reflected on the whole experience, it was hard to pin point the exact elements that caused a specific experience. Also time and participant exhaustion did prevent going into detail with user experience aspects and conducting a laddering interview. We would therefore recommend an evaluation where participants can explore the system freely in combination with valence method and a laddering interview.

The valence method by [Bu10] is a formative evaluation method to measure the user experience. While freely interacting with a product, the users label the moments in which they have positive or negative feelings with so called "valence markers". In a retrospective interview you go through each marker with the participant. The laddering technique is used to determine the need that lies beneath the experience in that moment, by questioning the users more precisely with every question about the design aspect and the experience at each marker. Incorporating this method evaluates what design element causes what type of experience and which basic human need matches.

"Talk about each basic human need individually and in depth."

If you do not have a fully functioning system or lacking the time and expertise to conduct an evaluation using valence method with a laddering interview another approach could be to simply talk about the basic human needs individually. For every basic human need ask the participants if they see that this would be addressed by the system and how. Of course you cannot for sure say if the need really would be addresses or not, but you get an insight of what your users are anticipating. For this kind of evaluation you only would need a low-fidelity prototype or mock-ups to visualize the scenario for the participants.

"Find the right mix of help positions and autonomy."

This is kind of hard. And probably needs more than one pre-test to see if the amount of help positions and autonomy work well together. If balanced out correctly and considering your specific user group and context of use, this will lead to better and more valid results.

A first evaluation of the concept with a low-fidelity prototype provided basically no help positions and no autonomy as there was only one predefined way to click through it. Despite that, participants showed signs of being cognitive overloaded (loosing orientation, poor concentration, not being sure anymore what the task was). The ux evaluation provided a more detailed introduction before interacting with the system, by putting the participant into a real life scenario. During task execution, participants got no further help neither from the system nor the moderator. Participants could freely explore the different items the prototype provided, rather than giving them just one choice. Although having a more complex and difficult task, participants showed no signs of cognitive overload. We explain that with the right mix of introduction and help positions, but still giving the participants an autonomous feeling through the concept. We concluded therefor that it is important to have the right mix of autonomy and help positions in an ux evaluation to not overstrain users but also to not give too much away, so a positive experience can develop. But we also emphasize that this fact is not necessarily generalizable. To generalize this, it would take more evaluations like this.

4 Related work

No recent contributions on (usability) evaluations of policy authoring tools are aware to us. Following, we list the contributions that had an impact on our evaluation or can be seen as similar.

[La17] evaluated in a use case study if the classical usability evaluation approach in a laboratory setting is applicable to ux evaluations. They concluded that the setting has a huge impact on participants' perceived ux. The authors talk in depth about their findings and reasons for their results, as well as providing possible alternatives and tips for future ux studies. The subject of their evaluation was the online shopping platform amazon and a digital camera. As these are already full functioning services that people already know and use out of intrinsic motivation, we do not feel confident comparing that study with our evaluation. Future work needs to show the applicability to an area that most users only use because they have to, rather than being intrinsically motivated.

[Th18] addresses the urgent need for privacy protection tools regarding personal data collected by wearable devices and smartphones. They request a framework that enables users to determine what kind of data may be collected and processed by the device. To be able to control that, users have to be able to create access control policies. The authors are planning on developing a policy authoring tool with a user-friendly interface by including interviews, surveys and laboratory experiments in the development process. Beyond usability they also want to test the correct understanding of the users' policy specifications. To the best of our knowledge no testing has been conducted so far, as the paper has been published recently.

The authors of [Ka06] included as a ux measurement a questionnaire on users' satisfaction on the quality of their policies into their usability evaluation of a privacy policy prototype.

This is a good first approach on measuring ux, but as they did not further question the users until they knew what caused the results, it can only be seen as a superficial user experience evaluation.

5 Conclusion & Future Work

Conducting a ux study in a laboratory we recommend getting a highly functioning prototype or an almost finalized system. We would also recommend having several ux evaluations in different product stages but also to think about alternatives to laboratory evaluations (see also [La17]. When conducting a laboratory evaluation we recommend using the valence method combined with laddering interviews to get valid results about what basic human needs are addressed through what.

Using the UNeeQ right before and directly after the tasks provided good results. To be able to make a decent statement on how using the tool influenced the users´ experiences, one has to identify the state of mind the participants started the test with by letting them fill in the UNeeQ before the interaction with the tool. Comparing those results with the results from the UNeeQ after interaction, created more precise data. Although we did not get to know the underlying basic human needs. Nevertheless it gave insight how the interaction changed participants perceived ux.

One of the most important aspects we learned is: consider designing for ux and include adequate ux evaluations from the very beginning even though usability might still seem to be the most important factor.

6 ACKNOWLEDGEMENTS

This research is part of the LIGHT^{est} project funded from the European Union's Horizon 2020 research and innovation programme under G.A. No 700321

Bibliography

[Hu10] 9241-210:2010(en), I. Ergonomics of human-system interaction - Part 210: Human-centred design for itneractive systems, 2010.

[Bo06] Bonatti, P. A.: Flexible and Usable Policies. In: W3C Workshop on Languages for Privacy Policy Negotiation & Semantics Driven Enforcement in REWERSE, pp. 1–5, 2006.

[Br06] Brodie, C. A.: An Empirical Study of Natural Language Parsing of Privacy Policy Rules Using the SPARCLE Policy Workbench. In: Public Policy, pp. 8–19, 2006.

[Bu10] Burmester, M. et.al.: Valence method for formative evaluation of user experience Proceedings of the 8th ACM Conference on Designing Interactive Systems. In: Valence

method for formative evaluation of user experience Proceedings of the 8th ACM Conference on Designing Interactive Systems, pp. 364–367, 2010.

[CI06] Cao, X.; Iverson, L.: Intentional access management: making access control usable for end-users. In: SOUPS Proceedings of the second symposium on Usable privacy and security, pp. 20–31.

[KBK05] Karat, C.-M.; Brodie, C.; Karat, J.: Usability Design and Evaluation for Privacy and Security Solutions. In: Security and Usability: Designing Secure Systems That People Can Use, pp. 47–74, 2005.

[CG04] Cranor, L. F.; Garfinkel, S.: Secure or usable?. In: IEEE Security and Privacy, pp. 16–18. 2004.

[FRR09] Ferreira, A.; Rusu, C.; Roncagliolo, S.: Usability and security patterns. In: Proceedings of the 2nd International Conferences on Advances in Computer-Human Interactions, ACHI 2009, pp. 301–305. 2009.

[FI10] Fischer-Hübner, S.; Iacono, L.; Möller, S.: Usable Security und Privacy. In: Datenschutz und Datensicherheit - DuD, pp. 773–782, 2010.

[FP14] Fronemann, N.; Peissner, M.: User experience concept exploration: user needs as a source for innovation. In: Proceedings of the 8th Nordic Conference on Human-Computer Interaction Fun, Fast, Foundational - NordiCHI ’14. pp. 727–736, 2014.

[Ia18] Iacono, L. L. et.al.: Consolidating Principles and Patterns for Human-centred Usable Security Research and Development, In: European Workshop on Usable Security, London, 2018.

[Ka06] Karat, C.-M. et.al.: Evaluating interfaces for privacy policy rule authoring. In: Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI ’06, p. 83, 2006.

[KS14] Kirlappos, I.; Sasse, M. A.: What Usable Security Really Means : Trusting and Engaging Users. In: Human Aspects of Information Security, Privacy, and Trust HAS. Lecture Notes in Computer Science, p. 11, 2014.

[La17] Lallemand, C.: Lab Testing Beyond Usability : Challenges and Recommendations for Assessing User Experiences, pp. 133–154, 2017.

[MJ08] Meland, P. H.; Jensen, J.: Secure Software Design in Practice. In: 2008 Third International Conference on Availability, Reliability and Security, pp. 1164–1171, 2008.

[Pr17] Prieto, L. P. et.al.: Maybe poor Jhonny Really Cannot Encrypt - The Case for a Complexity Theory for Usable Security. In: CEUR Workshop Proceedings, pp. 53–59, 2017.

[Th18] Thuraisingham, B. et.al.: Towards a privacy-aware quantified self data management framework. In: 23rd ACM Symposium on Access Control Models and Technologies, SACMAT 2018, pp. 173–184, 2018.

[Vo17] Voronkov, A. et.al.: Systematic Literature Review on Usability of Firewall Configuration. In: ACM Computing Surveys, pp. 1–35, 2017.

[ZSS96] Zurko, M. E.; Simon, R. T.; Street, S.: User-Centered Security, pp. 1–9, 1996.