Perception and Consideration of the Explainees' Needs for Satisfying Explanations

Michael Erol Schaffer[1,*], Lutz Terfloth[1], Carsten Schulte[1] and Heike M. Buhl[1]

[1]Paderborn University, Paderborn, Germany

Abstract

To tailor explanations to individual explainees, explainers consider the explainees' developing knowledge and interests. For that, it is necessary that explainers monitor the behavior and utterances of the explainees. XAI should be able to perceive and react to explainees' needs in a similar manner to generate customized explanations. For this, it is a precondition to know which explanation needs are perceived by explainers and how explainers consider them in the explanation. With the goal to improve XAI-explanations in the long run, we investigated explanations of a less complex technological artifacts in a qualitative observation and interview study as a fist step. The research questions addressed explainers' perceptions of explainees' knowledge and interests. According to the dual nature theory, we differentiate between two distinct perspectives on technological artifacts: observable and measurable features addressing "Architecture" and interpretable aspects addressing "Relevance". Explainees can demand both duality sides and therefore, both sides should be addressed in the explanation. This became evident in our study. Hence, we discuss how our findings can be transferred to adaptive explainable systems.

Keywords

Mental Representations, User Model, Technological Artifacts, Human Explanations, Qualitative Analysis

1. Introduction

XAI would benefit from increased consideration of users, thereby implying that XAI should recognize these individuals with their needs [19, 17] as every user uses XAI in different contexts with different foci. Enabling XAI to adapt to the specific needs of users allows customized explanations [11]. User-centered XAI should aim to answer the following questions [16]: What needs to be explained? How does it need to be explained? And who does it need to be explained to? For this, XAI should have a user model that contains information regarding users' developing knowledge and interests. To create such a user model, it has to be understood how explainers' mental representations of the interlocutors evolve, which information they contain and which aspects need close monitoring [2, 3]. Therefore, in this phase of our empirical study, we investigated everyday explanations, in which explainers, EX, gave the explanation and

Late-breaking work, Demos and Doctoral Consortium, colocated with The 2nd World Conference on eXplainable Artificial Intelligence: July 17–19, 2024, Valletta, Malta

*Corresponding author.

michael.schaffer@uni-paderborn.de (M. E. Schaffer); lutz.terfloth@uni-paderborn.de (L. Terfloth); carsten.schulte@uni-paderborn.de (C. Schulte); heike.buhl@uni-paderborn.de (H. M. Buhl)

0009-0001-5821-9967 (M. E. Schaffer); 0000-0003-1134-5090 (L. Terfloth); 0000-0002-3009-4904 (C. Schulte); 0000-0002-1001-492X (H. M. Buhl)

explainees, EE, received the explanation regarding the explanandum [7], which was a simple technological artifact. We aimed to answer the following research questions:

- (RQ1) Which aspects of the EEs' knowledge of the technological artifact do EXs perceive during an explanation?
- (RQ2) Which aspects of the EEs' interests in the technological artifact do EXs perceive during an explanation?

2. Related Wok

2.1. Technological Artifacts and the Dual Nature Theory

Technological artifacts, like XAI, are made by humans to fulfill certain purposes [12, 28]. According to the philosophy of technology they possess a dual nature—comprising an Architecture, AR, and a Relevance, RE, side [12, 22, 27]. And potentially, this duality could deliver a structure for synthesis of explanations. AR incorporates observable features like structures and codes, and resembles an objective perspective. In contrast, RE refers to purposes and intentions, thereby resembling a subjective perspective. If this theory can be applied in every explanation of technological or even digital artifacts needs further investigation. But generally, when explaining technological artifacts, both duality sides should be addressed, if the goal of the explanation is understanding [21, 25]. Therefore, we investigated what EXs perceive of EEs in terms of the dual nature of technological artifacts. From a co-constructive perspective, EXs and EEs develop the explanation conjointly [18, 24]. The EXs monitor what the EE knows and wants to know regarding the technological artifact and adapt the explanations toward perceived needs of EEs [18]. If the assumption regarding the EEs' interests and knowledge are not accurate, the explanation might not be satisfactory, as the focus could erroneously be on AR when RE is required or vice versa [21, 25]. The importance of considering knowledge and interests in explanations has been stressed before [24, 29]. Therefore, we consider it necessary for XAI to possess a user model that incorporates developing knowledge and interests of users, too.

2.2. Knowledge

The EXs possess mental representations of the EEs' increasing knowledge [4]. This mental representation is structured by the duality of technical artifacts and changes throughout the explanation [8, 10]. The following questions illustrate what the knowledge representation is embracing: "What does the EE already know about a specific domain with regard to the dual nature of technological artifacts?" or "Which knowledge does the EE still need to acquire with regard to the dual nature of technological artifacts?"

2.3. Interests

The EXs also have mental representations of interests of EEs [9]. The EXs need to know what the EEs want to know and what they are interested in. Interests have an impact on the direction an explanation takes. As EXs monitor EEs' behaviour during the explanation, it is important to know what EXs perceive of EEs interests. EXs' perception of EEs' interests is based on inferring

in the situational context, which we refer to as "interests", and on directly expressed interests through questions or comments by EEs, which we refer to as "expressed interests" [23].

3. Which Needs Should XAI Perceive and Consider?

The EXs need to know how to explain the technological artifact. Therefore, having a representation of what EEs are interested in, which outcome they expect [17] and which knowledge can be built on is mandatory in explanations. The following model (see Fig. 1) could serve as a framework for explanation synthesis. Its development will be described in the results section. This model considers developing knowledge (which knowledge does the user have and which

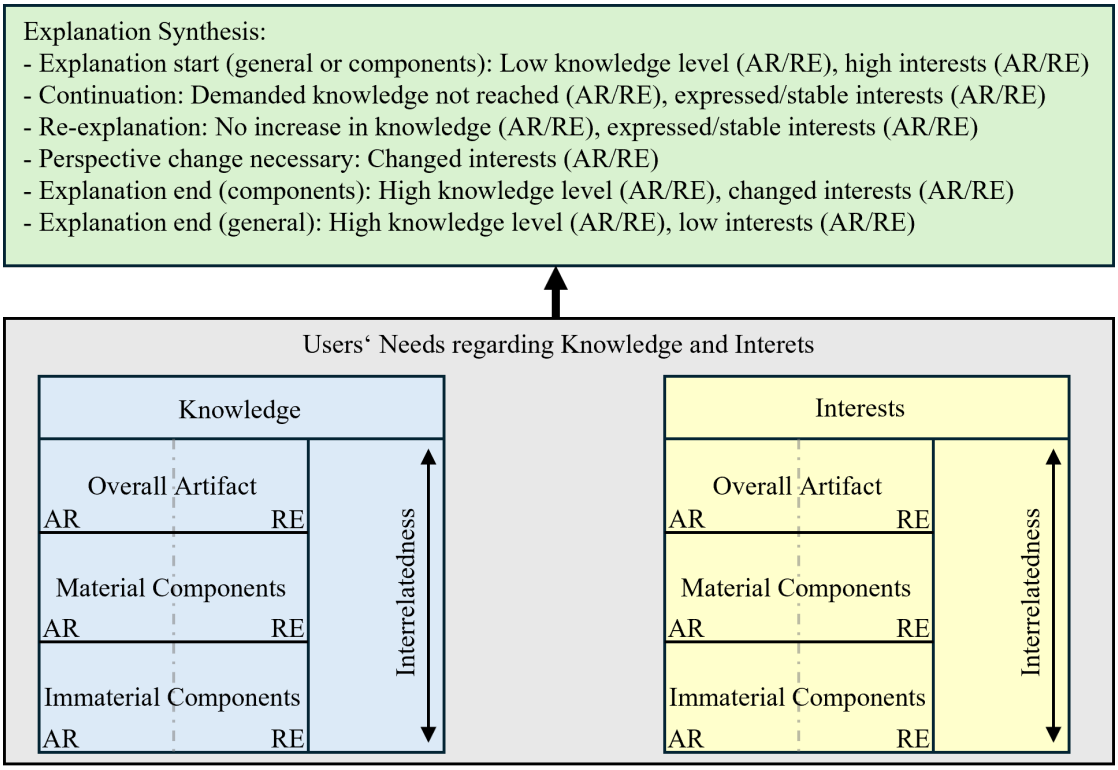

Figure 1: Users' Needs for Synthesis of Explanations (AR=Architecture, RE=Relevance)

knowledge is missing?) and interests (which interests does the user have in the technological artifact?) of users. Knowledge and interests relate to the overall artifact, material and immaterial components as well as their interrelatedness in regard to the dual nature .

A starting point for an explanation could be marked by high interests and low knowledge regarding the artifact [1]. It could be assumed to be an indicator for continuation of the explanation, when the demanded level of knowledge is not yet reached or interests regarding that aspect were expressed. When interests change, the foci or perspective of explanation might need to be changed as well. Moreover, a re-explanation of certain aspects might be necessary when knowledge does not increase but interests remain unchanged. The explanation could

potentially be stopped when no interests are expressed and satisfying levels of knowledge were reached. With this background, we deliver ideas for discussion and to support a socio-technological approach for XAI [14, 19]: We also argue that the dual nature regarding knowledge and interests should be part of the XAI's user model to generate satisfactory explanations that meet users' needs on both the AR and RE sides.

4. Method

In the context of naturalistic dyadic explanations, we assessed the EXs' perceptions of EEs' knowledge of and interests in the technological artifact. We followed a qualitative approach and used the qualitative content analyses, in order to explore the topic and related phenomena in-depth. Interviewees had the opportunity to give rich and meaningful answers.

4.1. Participants, Procedure and Material

EXs and EEs, were recruited (on-site and online) for naturalistic explanations. EXs were asked to familiarize themselves with the explanandum beforehand. We investigated nine explanations (*N=9*). All participants were students, aged between 21 and 32 years (*M=24.22, SD=3.46*). Our explanandum and technological artifact that needed to be explained is the strategic board game Quarto. In the explanatory process of Quarto, the dual nature unfolds as EXs and EEs need to address both AR and RE. This is similar to explanations of XAI, where both perspectives would be needed for understanding. We considered it reasonable to use Quarto in a first step before delving into XAI, which has higher complexity.

The study had three stages: pre-interview, explanation, and post-interview with video recall-interviews. Solely EXs were interviewed. The questions of the semi-structured interviews addressed the EXs' mental representations of EEs' knowledge and interests. After the pre-interview, which was conducted prior to meeting the EE, the EX was asked to explain the game to the EE. After the explanation, a post-interview was conducted. Retrospective video recall-interviews [6, 13] were conducted to reassess what EXs perceived in regard to EEs' knowledge of and interests in the technological artifact and its dual nature in specific explanation moments. For video recall-interviews, pre-selected sequences at the start, middle and end of explanations were shown and allowed the EX to report elaborately.

4.2. Content Analysis of Semi-Structured Interviews

The interviews were transcribed using standard orthography [15]. The deductive coding manual provided an overview of the characteristics of knowledge, inferred and directly expressed interests as well as AR/RE. Typical examples from interviews (see Table 1) were included. Segments ranged from single words to whole sentences but contained one specific aspect [20]. Segments were coded with knowledge or interest categories addressing AR or RE. We determined the intercoder reliability between two coders and Cohen's Kappa was k=.75, indicating an excellent agreement [5].

Table 1
Code System with Typical Examples.

Categories	**Typical Examples** (VP: Number of Study)
Knowledge	
Quarto AR	By now she just knew the rules.(VP16,VR3,Pos.7)
Quarto RE	She needed my experience on how to recognize situations.(VP26,VR5,Pos.17)
Board games AR	Everyone has in such a board game own pieces and colors.(VP24,Pre,Pos.38)
Board games RE	He knows how to develop personal, cooperative strategies.(VP20,Pre,Pos.31)
Interests	
Interests AR	Can you also build a diagonal row?(VP17,VR2,Pos.7)
Expressed Interests AR	Do I choose a piece for you? Or do I choose for myself?(VP26,VR3,Pos.5)
Interests RE	That you learn something in a game. A strategy.(VP20,Pre,Pos.39)
Expressed Interests RE	Why should I give you that piece so that you can win?(VP17,Post,Pos.11)

5. Results

For the model of how users' needs can be considered in explanations synthesized by XAI we bundled EXs' statements regarding EEs' knowledge and interests and identified main aspects that were important in order to answer the research questions. We then abstracted the concrete aspects of Quarto to generalize findings (see Fig. 1). An overview of distribution of coded segments across categories and explanation phases can be seen in Table 2.

Table 2
Knowledge/Interests: Absolute Frequencies

Knowledge	Pre	VR-S (15 VR)	VR-M (24 VR)	VR-E (11 VR)	Post
Board Games AR	46	3	1	0	42
Board Games RE	35	1	0	0	31
Quarto AR	0	50	88	31	92
Quarto RE	0	12	32	14	55
Total	81	66	121	45	220
Interests	Pre	VR-S (15 VR)	VR-M (24 VR)	VR-E (11 VR)	Post
Interests AR	58	15	16	2	49
Expressed Interests AR	0	10	50	25	34
Interest RE	107	1	5	0	45
Expressed Interest RE	0	1	18	2	20
Total	165	27	89	29	148

VR: Amount of Video Recalls in Explanation Phases: S:Start; M:Middle; E:End

5.1. Main Findings

EXs updated their mental representation of EEs' needs through monitoring of what EEs knew, did not know and wanted to know regarding the technological artifact and its dual nature. In the beginning, basics of the technological artifact with a focus on AR, for example characteristics of components, were explained. Usually, EXs perceived EEs' signals of developing knowledge of AR aspects. In many cases, especially after the early explanation focused on AR, EXs struggled to anticipate many of the EEs' interests in RE. Then EXs reported on directly expressed RE needs of EEs through questions or statements, for example on complex or outstanding aspects. These

aspects were particularly important for a complete understanding and for using the artifact. Toward the end of explanations, EXs learned a lot about EEs' needs regarding knowledge and interests and concluded that EEs understood most aspects on both the AR and RE sides. However, EXs believed that there were knowledge gaps and details that remained unclear. Usually EXs regarded the missing AR details as not important for a potential application of knowledge and therefore, were not explained in a closing manner. The missing RE knowledge though, targeted the interrelatedness of components and how to apply gained knowledge. Despite the fact that not every single aspect was understood to the furthest possible extent, EXs generally had the feeling that no further explanation was required due to the fact that a) not all aspects are equally important and b) not all aspects can be explained completely satisfactorily in an abstract manner and that gained knowledge needs to be tested or experimented with.

6. Discussion

To be potentially able to consider users' needs in XAI it is necessary to monitor [2, 3] multimodal behavior, questions and statements of EEs. Aspects that need to be monitored are the developing knowledge and interests with regard to the dual nature of technological artifacts [12]. We followed a qualitative approach to gain rich data to truly understand our research topic. We stopped conducting studies after we learned how the EXs' mental representations of EEs' knowledge and interests developed. The rather small sample size was sufficient for the development of the model and to continue our research with digital artefacts in the future. Our proposed model (see Fig. 1) on how users' needs can be considered in explanations synthesized by XAI could perspectively serve as a strategy for explaining concrete technological artifacts [11]. The model incorporates the dual nature theory. EXs perceived that EEs needed alternation between perspectives in a) aspects of higher complexity b) novel aspects c) unexpected aspects. If EXs would not have been aware of EEs interests, the ongoing explanation probably would have been less satisfactory for EEs [29]. Therefore, perspectives on the technological artifact were altering at different points in time and adapting to the needs of EEs. Both perspectives—AR and RE—were important for EEs to understand: a) EXs perceived different needs of EEs at different points in time regarding the artifact, and b) each artifact has its own unique features that might be challenging to understand for each individual EE. Furthermore, even if XAI is aware of the possibility of remaining knowledge gaps, there might be reasons for stopping the explanation. Because not every aspect needs to be completely understood, considering the potential context and extent of application. Alternatively, it might be beneficial to enhance explanations with visual elements [26]or to allow the interaction with the technological artifact to consolidate knowledge and practical application.

6.1. Future Work and Conclusion

The findings of this study provide hints on implications and further research. To increase comparability, a quantitative research approach with a bigger sample size is planned. We also aim to switch from analog technological artifacts to more complex digital artifacts. Even though we assume that our findings can be transferred to digital artifacts, we aim to to validate and extend our model for XAI by including concrete aspects of digital artifacts and their (im-)material

components. Again, the dual nature of technological artifacts might be useful as EEs can freely accentuate which features they are interested in and from which perspective they demand an explanation. We also plan to investigate how EEs' knowledge and interests develop and if their needs actually were met. A switch from naturalistic explanations to an experimental research design, where aspects of the interaction in regard to AR and RE are varied, is intended. We showed empirically how the EXs' assumptions about the EEs' knowledge and interests develop on multiple levels. The preliminary findings are important for the XAI context as knowledge and interests played a critical role in the explanations. To not consider EEs and their needs in explanations should be avoided by humans as well as XAI. Hence, knowledge and interests could serve as a base for the development of user models for person-specific and adaptive explainable systems. First ideas for XAI to synthesize explanations were provided.

Acknowledgments

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): TRR 318/1 2021 - 438445824.

References

[1] Buschmeier, H., et al.: Forms of Understanding of XAI-Explanations, http://arxiv.org/abs/2311.08760, (2023)
[2] Chi, M., Siler, S., Jeong, H.: Can tutors monitor students' understanding accurately? COGNITION INSTRUCT. 22, (2004).
[3] Clark, H.H., Krych, M.A.: Speaking while monitoring addressees for understanding. J. M. L. 50, 62–81 (2004).
[4] de Jong, T., Ferguson-Hessler, M.G.M.: Types and qualities of knowledge. Educ. Psychol. 31, 105–113 (1996).
[5] Fleiss, J.L.: Statistical Methods for Rates and Proportions. Second Edition. Wiley, John and Sons, Incorporated, New York, N.Y. (1981)
[6] Fox-Turnball, W.: Autophotography: A means of stimulated recall for investigating technology education. In: Benson, C., Lunt, J. (eds.) International Handbook of Primary Technology Education, Rotterdam (2011)
[7] Garfinkel, Alan: Forms of explanation: Rethinking the questions in social theory. Yale University Press (1981)
[8] Greca, I.M., Moreira, M.A.: Mental models, conceptual models, and modelling. Int. J. Sci. Educ. **22**, 1–11 (2000).
[9] Harackiewicz, J.M., Smith, J.L., Priniski, S.J.: Interest matters: The importance of promoting interest in education. Policy Insights Behav. Brain Sci. 3, 220–227 (2016).
[10] Jones, N.A., Ross, H., Lynam, T., Perez, P., Leitch, A.: Mental models: An interdisciplinary synthesis of theory and methods. ES. 16, art46 (2011).
[11] Kaplan, S., Uusitalo, H., Lensu, L.: A Unified and Practical User-Centric Framework for Explainable Artificial Intelligence. KNOWL-BASED SYST. 283, (2024).

[12] Kroes, P.: Technological explanations: The Relation between structure and function of technological objects. Soc. Philos. Technol. Q. Electr. J. 3, 124–134 (1998).
[13] Lyle, J.: Stimulated Recall: A report on its use in naturalistic research. Br. Educ. Res. J. **29**, 861–878 (2003).
[14] Neerincx, M.A., van Der Waa, J., Kaptein, F., van Diggelen, J., J.: Using perceptual and cognitive explanations for enhanced human-agent team performance. In: Harris, D. (ed.) Engineering Psychology and Cognitive Ergonomics. pp. 204–214. Springer International Publishing, Cham (2018)
[15] O'Connell, D.C., Kowal, S.: Transcription systems for spoken discourse. In: D'hondt, S., Östman, J.-O., Verschueren, J. (eds.) The Pragmatics of Interaction. pp. 240–254. John Benjamins Publishing Company, Amsterdam (2009)
[16] Ribera, M., Lapedriza, A.: Can we do better explanations? A proposal of user-centered explainable AI. IUI Workshops. (2019)
[17] Riveiro, M., Thill, S.: "That's (not) the output I expected!" On the role of end user expectations in creating explanations of AI systems. Artificial Intelligence. 298, 103507 (2021).
[18] Rohlfing, K.J., et al.: Explanation as a social practice: Toward a conceptual framework for the social design of AI systems. IEEE Trans. Cogn. Dev. Syst. 13, 717–728 (2021).
[19] Schoonderwoerd, T.A.J., Jorritsma, W., Neerincx, M.A., van den Bosch, K.: Human-centered XAI: developing design patterns for explanations of clinical decision support systems. Int. J. Hum.-Comput. Stud. **154**, Article 102684 (2021).
[20] Schreier, M.: Qualitative content analysis in practice. SAGE Publications Ltd, London (2012)
[21] Schulte, C.: Duality reconstruction—teaching digital artifacts from a socio-technical perspective. In: Mittermeir, R.T., Sysło, M.M. (eds.) Informatics Education—Supporting Computational Thinking. pp. 110–121. Springer Berlin Heidelberg, Berlin, Heidelberg (2008)
[22] Schulte, C., Budde, L.: A framework for computing education: Hybrid interaction system: The need for a bigger picture in computing education. In: Proc. of the 18th Koli Calling Int. Conference on Computing Education Research. pp. 1–10. ACM, Koli Finland (2018)
[23] Silvia, P.J.: Expressed and Measured Vocational Interests: Distinctions and Definitions. J. Vocat. Behav. **59**, 382–393 (2001).
[24] Sokol, K., Flach, P.: One explanation does not fit all: The promise of interactive explanations for machine learning transparency. Künstl. Intell. 34, 235–250 (2020).
[25] Soloway, E.: Learning to program = learning to construct mechanisms and explanations. Commun. ACM.**29**, 850–858 (1986).
[26] Schrills, T., Franke, T.: Color for Characters - Effects of Visual Explanations of AI on Trust and Observability. In: Degen, H. und Reinerman-Jones, L. (hrsg.) Artificial Intelligence in HCI. S. 121–135. Springer International Publishing, Cham (2020)
[27] Terfloth, L., Schaffer, M., Buhl, H.M., Schulte, C.: Adding why to what? Analyses of an everyday explanation. In: Longo, L. (eds.) Explainable Artificial Intelligence. pp. 256–279. Springer Nature Switzerland, Cham (2023)
[28] Vermaas, P.E., Houkes, W.: Technical functions: A drawbridge between the intentional and structural natures of technical artefacts. Stud. Hist. Philos. Sci. Part A. 37, 5–18 (2006).
[29] Wittwer, J., Renkl, A.: How effective are instructional explanations in example-based learning? A meta-analytic review. Educ. Psychol. Rev. 22, 393–409 (2010).