Towards Self-explaining Intelligent Environments

Serge Autexier
German Research Center for Artificial Intelligence (DFKI)
Bremen, Germany
serge.autexier@dfki.de

Rolf Drechsler
German Research Center for Artificial Intelligence (DFKI)
& *University of Bremen*, Bremen, Germany
rolf.drechsler@dfki.de, drechsler@uni-bremen.de

***Abstract*—Intelligent environments and assistants increasingly pervade and shape all areas of our daily life. They are characterised by dynamically varying numbers of coexisting, ideally cooperating but at least coordinated assistants and devices and users. This makes it difficult not only for technically inexperienced individual users to understand what is happening around them. It also does not help to know that the overall behaviour is correct in any sense whatsoever. From Human-Computer-Interaction research it is well known that systems providing explanations of their decision can improve the user's trust in the systems. Illustrated by examples from our research on developing smart assistance in intelligent environments we motivate the need for self-explainability capabilities of intelligent environments. We shed a light on different aspects of self-explainability, motivate the need for explanations for the target user groups and provide criteria that are relevant to address for the design of self-explaining intelligent environments.**

***Index Terms*—intelligent environments, assistance processes, human-computer-interaction, explanations**

I. Introduction

Intelligent environments such as smart homes equipped with complex assistance processes beyond simple *if-then*-control rules are more and more becoming a reality in our daily life. They consist of a variety of different systems that range from rather primitive ones, like motion sensors or remotely controllable lights, to more sophisticated ones, like gesture recognition systems or autonomously driving robotic vehicles.

Realising intelligent behaviour in such environments is a major task as it has to go all the way up from protocols to communicate on a low signal level to sophisticated services recognising, mediating, and planing high level activities. As a result the behaviour of intelligent environments implemented in assistance processes is governed by the logic of the assistance, the programming style and the programs at the different levels of abstractions. From the user perspective, however, the assistance processes appear as a monolithic program and especially for users without technical skills the behaviour exhibited by the environment may be obscure and raises questions. Not being able to provide explanations or answers at the right level of abstraction that consistently fit with the observable behaviour from the user point of view may hamper the trust in the assistance processes and lower user acceptance.

From Human-Computer-Interaction research it is well known that systems providing explanations of their decision can improve the user's understanding and trust in the systems [1], [8], [13], [19]. Based on that we aim at methods and ideally tool support that allows to develop assistance processes equipped with adequate self-explanation capabilities, that are maintained together with the assistance processes. Moreover it should allow for checking respectively verifying that the explanations are indeed coherent with the assistance processes, because only then we can guarantee that the understanding of the assistance processes implicitly build up by the users from the explanations complies with the system behaviour.

In recent work [9] self-explainability was proposed as a general feature for next generation digital systems. In this paper we follow that avenue and work out the pecularities that arise when considering assistance processes in intelligent environments. The rest of the paper is organised as follows: In Section II we present the smart home lab BAALL which sets the context for our long-term research about intelligent environments and user-interaction. From that research we present in Section III two selected examples of assistance processes and situations that demand for explanations for the users. Finally, we derive in Section IV a set of criteria that we deem relevant for the development of self-explainable intelligent environments.

II. The Bremen Ambient Assisted Living Lab

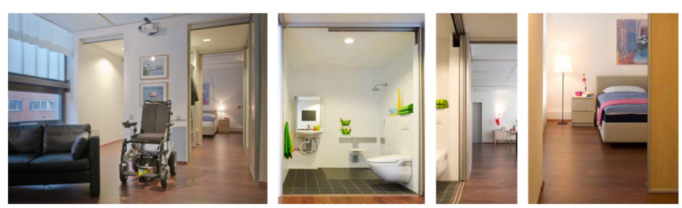

Fig. 1. The Bremen Ambient Assisted Living Lab

The research department Cyber-Physical Systems (CPS) of the German Research Centre for Artificial Intelligence in Bremen operates the *Bremen Ambient Assisted Living Lab (BAALL)*[1] (see Fig. 1) since 2009 [2], [10], [14], which is a 60m^2 smart home laboratory apartment in which assistance systems are developed according to the Design-For-All principle with the focus on automation, personalisation, intuitive human-technology interaction, and a special focus

The work presented in this paper results from the projects SHIP (Grant No. 01IW10002) and INTUITIV (Grant No. 16SV7979) funded by the German Ministry for Research and Education (BMBF)

[1]https://www-cps.hb.dfki.de/research/baall

on assistance systems for people with age-related limitations, especially mobility assistance.

More specifically, the research questions focus on how to actually enhance environments to serve as a basis for intelligent assistance by integrating IoT actors and sensors and how to functionally extend appliances. That research is guided by the ubiquitous computing [20] paradigm to make technology available where needed but as invisible as possible.

Based on that further research activities are devoted to the actual development of useful personal assistance processes, the mobility of assistance processes to follow users when they move around in their flat or leave home, the safety, security and reliability of assistance processes and finally the interaction between users and the assistance processes.

The research on the development of assistance processes is guided by the idea of *Companion* systems and aims for methods to design and deploy assistance systems, that adapt themselves to smart home environments and the users living in such environments and that provide their assistance in a flexible manner. The vision is to eventually allow for App Stores of assistance process for smart homes, similar to todays App Stores for smartphone applications, that everybody can download for his/her smart home and use it without the need for complicated configurations. The research questions are how to design assistance processes that reactively adapt themselves to local conditions, how to ensure a correct behaviour in the presence of multiple users and other assistance processes which concurrently interprete and control the smart environments. This is researched under the topic of modularity and compositionality properties for assistance processes.

With respect to human-machine interaction the target is to enable natural interactions between users and the environment, i.e. overcoming the need to use a smartphone or tablet to interact with the environment. The targeted natural forms of interaction are speech, gesture, haptic or vision and multi-modal combinations thereof. Based on that research is devoted to how to design multi-partner dialogues using multi-modal interactions and how to design interaction to ensure acceptance and trust by the users. This is where the present paper comes in to illustrate the different forms how explanations or intrinsic behavioural intuitiveness of assistance processes can serve to improve their acceptance and use.

The BAALL serves as a test and evaluation environment for research projects as well as for courses and theses at the University of Bremen and is a demonstration room for visitor groups from academia and industry as well as for the general public. As such it allows to gain feedback and insights from diverse groups of people, especially also target user groups for assistive technologies and typically without or little technical skills or even affinity. Moreover, it allows to place assistive technologies in a realistic context and analyse requirements and shortcomings of solutions. The examples presented in the following section result from requirement analysis for specific assistances, evaluations of assistances and methods as well as feedback and observations from user groups.

Fig. 2. Automatic Driving Wheelchair Rolland and Walker

III. Motivating Examples

In this section we provide two examples from own research in the context of intelligent environments, that illustrate the need for self-explainable intelligent environments, as well as different reasons giving rise to such a need and the means to provide explanations.

A. Automatic Driving Assistants

The BAALL group has a long research history[2] on developing smart mobility assistants for persons with physical and/or psychological impairments [12], [15]–[18]. The focus of this research is to start from physical helper systems such as wheelchair or walkers and to enhance these by sensors and actors and assistances to support activities of daily living. As an example consider the electric wheelchair *Rolland* on the left-hand side of Fig. 2, which, equipped by LIDAR sensors, provides assistance to avoid obstacles in manual driving mode. Furthermore, it provides automatic driving assistances using localisation in known indoor environments based on SLAM maps and outdoors using GPS information and map data from OpenStreetMap. The identical technology has also been integrated on a walker (right-hand side of Fig. 2) which has electric wheels and thus provides the same assistances for obstacles avoidance and automatic driving.

In automatic driving mode both systems can be given a target position and use a global path planner to find a route to reach the target position and then start driving towards it. During driving they continue scanning for (temporary) obstacles, e.g. persons walking around, and use a local path planner avoid them by trying to drive around them or, in case this is not possible, stop and wait until the path is free again.

The observation now is that persons who experience the automatic driving wheelchair or walker often are not at ease with the systems, as the movements of the systems resulting from the global and local path planners are not foreseeable for these persons. This makes it difficult for persons to behave normally in the presence of the systems, when the systems are in the same space, to know where the systems want to got and

[2]It started at the university of Bremen around 1997 and was then included in DFKI around 2006.

if they shall make space for them. Or, for instance in a corridor, if the systems move towards the persons, to know when they will start adapting their trajectory to avoid the persons and to which side.

The automatic wheelchair and walker are representative for a variety of automatic driving robots and a small version of what automatic driving cars are doing on a large scale. The difference to classical robot installations which are behind safety fences is that close contact is omnipresent in these systems for home use or nursing care. The problems sketched above in the close contact with humans result from that, up to now, little attention has been paid to the optimal design of communication between robot and persons, patient, nursing staff or third parties during encounters while driving automatically in the sense of the optimal choice of communication mode (non-verbal kinematic, non-verbal iconic, verbal). The communication must be closely integrated together with the robot design and there is a need to have a strong coherence between communicative acts and the robot's behaviour, in order to reach the goal of enabling persons to understand the robots intentions and thus improve their trust in the robot. The communicative acts are in this case a form a self-explainability of the robot system, which is provided via intuitiv kinematic movements, speech and iconic communication acts.

The project INTUITIV[3] (German Ministry for Research and Education, BMBF, Grant No. 16SV7979, June 2018-May 2021) addresses this problem and aims for developing intuitive non-verbal and informative verbal forms of communication between robots and humans, which can be transferred to diverse application domains of robots in the service sector with direct human-robot interaction. The project uses rehabilitation environments as application domain and studies how to integrate different modalities in order to achieve the best communication behaviour. For this purpose, methods are developed that combine visual and auditory perceptions to form coherent inputs and select and synchronise suitable output modalities. Solutions for the following problem areas are developed to realize the automatic movements for automatic robotic systems in the application environment: First, reliable and robust localisation and navigation to deal with variable numbers of people and occlusions. Secondly, driving models that result in movements that people can anticipate—possibly combined with verbal and iconic modalities. In order to use verbal communication, it investigates how relationships between humans and robots can be established, maintained and resolved appropriately. This includes, in particular, dialogues about intentions that are not clearly recognisable by movement patterns or iconic information.

B. Experiencing Smart Assistances

The second example illustrating the need for self-explainable intelligent environments comes from the research on the design of assistance processes that reactively adapt themselves to local conditions and to ensure a correct behaviour in the presence of multiple assistance processes.

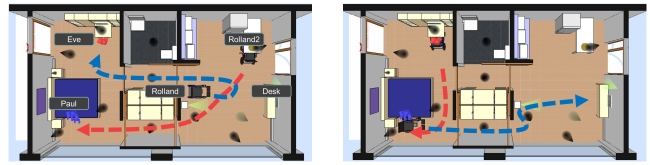

Fig. 3. Conflict Situations resolved by Scheduling

The general problem for the users experiencing multiple assistance processes is that the devices in the environment may perform an action and the reason for it is unknown to the users. Or a request from the user to the assistance processes is not addressed and the reason is, again, unknown to the user. In either case there is a need to provide an explanation for the action visible to the user or why no action is performed (see also [13]). The reason depends on the logic implemented in each assistance process and the current situation of the environment and may also be an emergent behaviour from the presence of multiple assistance processes.

To illustrate this, we consider the following two example assistance processes that have been realised in the BAALL as part of the research project SHIP[4] (*Semantic Integration of Heterogenous Processes*, German Ministry for Research and Education, Grant No. 01IW10002, January 2011-December 2013), see [4], [5]. The SHIP project was concerned with developing formalisms and tools to model multiple adaptive, reactive and flexible assistance processes by monitoring and orchestrating the different devices in a smart home, ranging from rather primitive ones, like motion sensors or remotely controllable lights, to more sophisticated ones, like gesture recognition systems or autonomously driving robotic vehicles. To this end the SHIP project developed a concurrent programming language representing states in Description Logics [11] and state transitions as logical updates [3] enabling deductive support to infer non-explicitly represented knowledge. It uses temporal logic to suspend execution of a process for a particular evolution of the global state that is specified by a LTL formula [6], [7]. Since a process can fork into subprocesses this provides a mechanism for runtime verification by splitting a process into a subprocess executing some critical program and another parallel subprocess monitoring the first one by waiting for the desired evolution of states specified in its LTL formula.

The two illustrating assistance processes realised using the SHIP language are:

a) Night Surveillance Assistance: The first example assistance is a night surveillance service, where at night time, doors are automatically closed. If persons move around in the flat and open doors, the doors are automatically closed again after a short delay. In case of an emergency such as a fire alarm, doors are automatically opened and the whole flat is illuminated.

[3]https://www-cps.hb.dfki.de/research/projects/INTUITIV

[4]https://www-cps.hb.dfki.de/research/projects/SHIP

b) Transportation Assistance: The second assistance is a transportation assistance for the persons in the BAALL managing the differengt available automatic wheelchairs (see Section III-A). In addition to managing the transportation requests from persons and wheelchair rides, the service also opens closed doors if a wheelchair has to drive through it. For the comfort of the residents, the service also takes care that when wheelchairs drive autonomously at night time in rooms where persons are present, the lights are turned on.

Each individual assistance processes is based on a model of the situation and involved recursive and concurrent subprocesses to exhibit the desired overall behaviour. The processes adapt themselves to the local situations using ontology-based configurations describing the available devices and smart home topology. The night surveillance is only active in night mode and that information is communicated to the persons using a specific RGB lamp set to a blue color at night time. This allows the user to *understand*, why a door is closed or a light turned off because the flat is in night mode. Obviously, only the user informed about this behaviour of the assistance process is able to come to the right conclusion, while a novice user would first have to get an explanation of this behaviour on the right level of abstraction. This implies that explanations need not only be at the right level of abstraction and compliant to the observable program behaviour, but also that detailedness of explanation varies depending on the model a user has already build up (see also [8] for the role of user models for explanations).

In the transportation assistance several information need to be provided to the user. The first and simple ones, which are just feedbacks, is to inform if a request for a transportation is taken up, when a wheelchair will be coming etc. This is feedback to the user and not explanations in the sense under investigation in this paper. The more intriguing situations for the users—even when they are used to peculiarities of the automatic driving wheelchairs as described in the Section III-A—are when there are additional deviations form the expected behaviour due to the presence of other automatic driving wheelchairs. This may occur when several wheelchairs are driving and the overall ride control implemented as part of the transportation assistance requires a wheelchair to take a different route, which is not necessarily the shortest, or to stop and wait to let another wheelchair go first. This is depicted by the left-hand side illustration in Fig. 3, where the wheelchair routes are indicated by dashed lines with arrows. Or else, as depicted by the right-hand side illustration of Fig. 3, a wheelchair that is suddenly moving away because it is blocking the route for another wheelchair. In all these situations the users' understanding (and thus trust) of the system can be improved by providing an explanation for the reason of the observable behaviour. Of course, that explanation needs to be conveyed in a form that is intelligible for a user without technical knowledge. A complication to provide explanations in these situations results from the fact that the overall behaviour results from behaviours at different levels of abstraction: on the lower level, there are the global and local path planner of the wheelchairs, which need to be explained as isolated systems to gain understanding. As part of the teansportation scenario the wheelchairs are addionnally governed at a higher level by the ride request and scheduling transportation assistance. As the lower level has been developed before the higher level, there is a need for being able to obtain explanations by *hierarchical* composition of the explanations at the different levels.

Finally, if the night surveillance and the transportation assistance run simultaneously, there are situations, where doors are opened or lights turned although the flat is is in night mode, just because wheelchairs need to pass through. Hence there is need to obtain explanations for this behaviour by *horizontally* composing explanations from the individual assistances.

IV. Criteria for Self-Explaining Intelligent Environments

In this section we provide a set of criteria with questions to answer that we judge necessary to address when designing self-explainable assistance processes for intelligent environments. In the description we distinguish the *user level*, which is yet a vague notion describing the level at which normal, non-technically skilled users would describe or understand program behaviours, from *program level* as the level, at which assistance process behaviours are described, which can be a programming language, but can also be other more low level specifications of the behaviour. This is in line with the different types of explanations introduced in [9] supporting different perspectives and various levels of granularity:

- *User-level resp. user-understandable explanations* refer to input data and the user-visible conceptual state that triggers an action.
- Specification-defined explanations remove all references to input data and the system state to reduce the explanation to the specification relevant for certain actions.
- Architectural explanations show which modules contributed to triggering an action.
- *Program-level explanations* justify the execution paths taken by a program or hardware units.

A. Behaviour models at different semantic levels

In the context of intelligent environment, the target users typically have no technical skills and expertise and thus in general have semantic models, not necessarily algorithmic ones, which they use to develop an understanding of technical systems or intelligent environments. These models are likely very different from the algorithmic models at program-level. Thus there are two key players that need to be linked: the program and the user and generating explanations consists among others of aligning *behaviours at the program-level* and *behaviours at the user-level*. Furthermore, the user may perceive an intelligent environment as a single system and hence his behaviour model will be monolithic. However, the actual intelligent environment may consists of several components possibly acting concurrently and hence the behaviour model at program-level may already structurally be very different from the behaviour model at user-level.

B. Complete and partial behaviour models

At program-level typically a complete model of the system behaviour is available, if the source code is known. However, trying to obtain a complete system model at the user-level may be sufficient to provide the user an answer about how in general a system functions, but is inadequate to answer questions why specific events occurred. Here a partial model grounded to the specific situation allowing to provide an explanation is sufficient. Over time several explanations are provided each consisting of grounded partial models which will result in that the user will assemble and generalize the grounded partial models to semi-grounded integrated behaviour models. This model will enable the user to predict the behaviour of the environment and if this is consistent with the system behaviour, then it will improve trust in the system. However, if it does not, then the actual goal of providing explanations is not achieved. This also implies that the provision of partial models is deterministic, i.e. if the same situation happens again, then the same (or an equivalent) partial model should be generated. Hence, there is a need to ensure that the sequence of partial behaviour models are combined to more generic partial model that is compliant with the complete behaviour model. Moreover, the combination of the partial models should be unambiguous in the sense that all possible combinations result in partial models that are behaviourally equivalent.

C. Criteria for Self-Explainability

From the sketched examples and observations in the previous subsection, we derive the following set criteria with accompanying questions to answer that we deem relevant to be addressed for the design of self-explainability capabilities.

1) *Intuitiveness*: What are behaviour models that allow to describe behaviours in an intelligible manner for normal users, i.e. at user-level?
2) *Linkability*: Is there a way to link behaviour models at program-level with behaviour models at user-level?
3) *Compositionality*: How to handle the differences in the structuring of behaviour models at program-level and user-level?
4) *Definability*: If so, how can this be defined such that the behaviour models at user-level can be computed automatically from behaviour models at program-level?
5) How to obtain a behaviour model at user-level for systems, where the behaviour model at program-level is unknown or there is no way to define such a mapping (e.g., machine learning based systems)?
6) *Situatedness*: How to generate partial grounded models at user-level?
7) *Integration*: What are means to combine/integrate partial behaviour models, that correspond to how users implicitly combine models?
8) *Clarity*: How to generate partial grounded models constrained by past grounded models such that the combined partial model is unambiguous?
9) *Compliance*: How to verify/ensure that a partial model complies with the overall behaviour model?

V. Conclusion

By reviewing previous research on developing assistance systems and the reactions and feedback obtained from target user groups during evaluations and demonstration of the assistances, we presented two examples that are representative to motivate the need for self-explainability capabilities in intelligent environments. The examples illustrate the kind of reasons for actions or non-actions that need to be explained to users in an intelligible manner and the problems to overcome to obtain explanations from hierarchical and horizontally composed assistances processes. From these we have derived a set of criteria with associated questions that need to be addressed when engaging into research and developments on methods and tool support for the design of self-explainable intelligent environments.

Acknowledgment

We thank all our colleagues from the Research Department Cyber-Physical Systems of DFKI—especially Daniel Grosse, Dieter Hutter, Christoph Lüth and Christian Mandel—for discussions and feedback on the topic of self-explainability and intuitiveness of intelligent environments.

References

[1] Dympna O'Sullivan Adrian Bussone, Simone Stumpf. The role of explanations on trust and reliance in clinical decision support systems. In *Healthcare Informatics (ICHI)*, pages 160–169. IEEE, 2015.

[2] Serge Autexier, Rossitza Goleva, Nuno M. Garcia, Rumen Stainov, Ivan Ganchev, Constandinos X. Mavromoustakis, Ciprian Dobre, Ivan Chorbev, Vladimir Trajkovik, and Eftim Zdravevski. Enhanced living environments: From models to technologies. In Rossitza I. Goleva, Ivan Ganchev, Ciprian Dobre, Nuno M. Garcia, and Carlos Valderrama, editors, *Enhanced Living Environments: From models to technologies*, chapter End-users' AAL and ELE service scenarios in smart personal environments, pages 101–132. IET Digital Library, 2017.

[3] Serge Autexier and Dieter Hutter. Constructive DL update and reasoning for modeling and executing the orchestration of heterogenous processes. In *26th International Workshop on Description Logics (DL-13)*. CEUR Vol. 1014, 2013.

[4] Serge Autexier and Dieter Hutter. Ship - a logic-based language and tool to program smart environments. In Moreno Falaschi, editor, *Post-Proceedings 25th International Conference on Logic-Based Program Synthesis and Transformation (LOPSTR 2015)*, LNCS. Springer-Verlag, october 2015.

[5] Serge Autexier, Dieter Hutter, and Christoph Stahl. An implementation, execution and simulation platform for processes in heterogeneous smart environments. In Juan Carlos Augusto and Reiner Wichert, editors, *Fourth International Joint Conference on Ambient Intelligence*, LNCS, Dublin, Ireland, december 2013. Springer.

[6] Franz Baader, Silvio Ghilardi, and Carsten Lutz. LTL over description logic axioms. *ACM Transactions on Computational Logic*, 2012.

[7] Andreas Klaus Bauer and Yliès Falcone. Decentralised ltl monitoring. In Dimitra Giannakopoulou and Dominique Méry, editors, *FM 2012: Formal Methods - 18th International Symposium, Paris, France, August 27-31, 2012. Proceedings*, volume 7436 of *Lecture Notes in Computer Science*, pages 85–100. Springer, 2012.

[8] Anind K. DeyIn. Explanations in context-aware systems. In *ExaCt*, pages 84–93, 2009.

[9] Rolf Drechsler, Christoph Lüth, Goerschwin Fey, and Tim Güneysu. Towards self-explaining digital systems: A design methodology for the next generation. In *3rd International Verification and Security Workshop (IVSW), Costa Brava, Spain*, 2018.

[10] Jochen Frey, Christoph Stahl, Thomas Röfer, Bernd Krieg-Brückner, and Jan Alexandersson. The dfki competence center for ambient assisted living. In Boris de Ruyter, Reiner Wichert, David V. Keyson, Panos Markopoulos, Norbert Streitz, Monica Divitini, Nikolaos Georgantas, and Antonio Mana Gomez, editors, *Ambient Intelligence: First International Joint Conference*, volume 6439 of *Lecture Notes in Computer Science*, pages 310–314. Springer, 2010.

[11] Ian Horrocks, Oliver Kutz, and Ulrike Sattler. The even more irresistible SROIQ. In *Knowledge Representation KR2006*. AAAI Press, 2006.

[12] Jörg Kollmann, Axel Lankenau, Andreas Bühlmeier, Bernd Krieg-Brückner, and Thomas Röfer. Navigation of a kinematically restricted wheelchair by the parti-game algorithm. In *Spatial Reasoning in Mobile Robots and Animals, AISB-97 Workshop*, pages 35–44. Manchester University, 1997.

[13] Brian Y. Lim, Anind K. Dey, and Daniel Avrahami. Why and why not explanations improve the intelligibility of context-aware intelligent systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, page 2119–2128. ACM, 2009.

[14] Christian Mandel and Serge Autexier. People tracking in ambient assisted living environments using low-cost thermal image cameras. In *Proceedings 14th International Conference on Smart Homes and Health Telematics (ICOST 2016)*, LNCS, Wuhan, China, june 2016. Springer-Verlag.

[15] Christian Mandel, Ahmit Choudhury, Serge Autexier, Karin Hochbaum, Jeannine Budelmann, and Petra Wedler. Gait cycle classification for wheeled walker users by matching time series of distance measurements. In *Proceedings Rehabilitation Engineering and Assistive Technology Society of North America (RESNA)*, july 2018.

[16] Christian Mandel, Tim Laue, and Serge Autexier. Smart wheelchairs and brain-computer interfaces. In Pablo Federico Diez, editor, *Smart Wheelchairs and Brain-computer Interfaces*, chapter Smart Wheelchairs. Elsevier, may 2018.

[17] Christian Mandel, Thomas Röfer, and Udo Frese. Applying a 3dof orientation tracker as a human-robot interface for autonomous wheelchairs. In *Proceedings of the 10th International Conference on Rehabilitation Robotics*. IEEE, 2007.

[18] Thomas Röfer, Christian Mandel, Axel Lankenau, Bernd Gersdorf, and Udo Frese. 15 years of rolland. In *Festschrift Dedicated to Bernd Krieg-Brückner on the Occasion of his 60th Birthday*, pages 255–272. Sichere Kognitive Systeme, DFKI Bremen, 2009.

[19] Margaret Burnett Sherry Yang Irwin Kwan Weng-Keen Wong Todd Kulesza, Simone Stumpf. Too much, too little, or just right? ways explanations impact end users' mental models. In *IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 2013.

[20] Mark Weiser. The computer for the 21st century. *Scientific American*, 265(3):94–104, 1991.