Towards Privacy-preserving Personalized Social Robots By Enabling Dynamic Boundary Management

Manuel Dietrich
Honda Research Institute Europe
Offenbach am Main, Germany
manuel.dietrich@honda-ri.de

***Abstract*—Designing personalized social robots, which should become an intrinsic part of everyday life, raises new challenges on how to respect the privacy of people interacting with them. In this paper, we introduce a little recognized conceptual perspective on privacy which, in our opinion, is highly relevant for the design of personalized social robots. This conceptual perspective, introduced by Irwin Altman [1] highlights that in everyday social life, especially with respect to interpersonal interaction, privacy preferences are not static. Instead, they tend to be in constant change dependent on context and experience. What matters for privacy is the possibility to dynamically reconfigure the degree of disclosure of personal information – in Altman's words being able to continuously manage the boundaries. Since personalized robots, for instance in their role as companions or life assistants, can be seen as artificial social actors Altman's perspective becomes relevant for their design. Taking the conceptual thoughts into account leads to a privacy-design strategy, which discards common approaches of an a priori configuration of preferences and rather targets a selective disclosure of personal information as part of a continuous boundary management process. In this paper, we discuss how we apply this perspective into a design process for robots, which are able to maintain long-term interaction in a privacy-preserving way.**

I. Introduction

Robots are about to enter many spheres of everyday life, for instance in the role as coworkers, companions, servants or life assistants. For all applications, they have to be able to adapt to spatial circumstances as well as to the specifics of the social situations they are thrown in. Since it is the mission of social robots to stay for longer, they have to be adaptive and flexible as well as technically robust, trustful and reliable. For an efficient assistance or companionship, it is central that robots learn to know people better and adapt to their characteristics and preferences. For many of the required capabilities of social robots, it is necessary that personal information is gathered and processed. Since the usage of personal information is intrinsic for the functionality, in designing such systems we have to take care of privacy.

Designing technical systems in a privacy-preserving way has multiple dimensions whereby security aspects as well as legal requirements have been discussed most (see next section). One direction, which is discussed less, is to account for privacy as a social-psychological phenomenon.[1] One key aspect of this perspective is the dynamic nature of privacy preferences with respect to everyday face-to-face interpersonal interaction and consequently the structural requirements of being able to manage these. We will transfer this perspective into the domain of personalized robotics.

Next, we will briefly map existing approaches of privacy-preserving design and locate our approach within the landscape. Following, we present the theoretical framework, which we argue to be relevant for the design of personalized social robotic systems. In a design for privacy-section, we sketch how the theoretical findings could be embedded into design.

II. Related Work: Designing for Privacy

In many regions of the world, guidelines and principles are developed which should align the design of personal data processing in interactive systems. Data protection regulations oblige to inform users about the ongoing or planned data handling practices whereby the information should contain the specification of the legitimate purpose of usage. Furthermore regulations oblige to provide documentation for demonstrating compliance (accountability), to allow users meaningful choices (consent) and to implement sufficient security measures (see for example: [5], pp. 22-24). Such or similar rules can be found in the General Data Protection Regulation [6], the European Union has recently enforced or in the older OECD as well as the US based FIP guideline.

In the privacy-engineering research we can observe a clear focus on the development of methods and architectures which target to improve secure data handling, for instance ways to protect the exchange of personal data without third party access (encryption, data storage infrastructure) or to make sensitive data public or available for third parties without the possibility to link the data to individuals (differential privacy, pseudo-anonymization techniques).[2] Other parts of privacy-engineering are technical methods which allow to reduce the amount of personal data processed while using a

[1] In the research field of Ubiquitous Computing are some approaches who have considered this theory for design: [2], [3], [4]

[2] Spiekermann and Cranor present a framework which states the relation between the degree of privacy users have with respect to the linkability of data to personal identifiers, [7], p. 75.

service or computing system. For instance a telepresence robot which is automatically blurring details of its video stream, for instance family photos on the background wall, so that persons operating the robot cannot see them [8]. These measures, especially those which prevent security breakdowns are one backbone of information privacy. Admittedly, research on new security and anonymization techniques is important, but there are other aspects relevant for privacy that are researched less. In their privacy-by-design position paper, Rubinstein and Good criticize the focus on "back-end security engineering": They claim that "several of the privacy incidents" are rather "nuanced violations of users' perception of privacy and their [users] choices regarding the context" ([9], p. 1352). For this the other facets of persevering privacy with respect to a technical system have to be discussed more intensively. The question is what it means to have the *right* perception and a context-adequate choice. We argue that the theoretical framework, which is presented in the next section, can help to answer this question.

III. The Dynamics of Privacy Preferences: Privacy as Boundary Management

In this paper, we refer to a theoretical direction which claims that humans' privacy preferences, especially with respect to sharing information in human-human interaction, are not static rather dynamic.

The social-psychologist Irwin Altman had first discussed this perspective on privacy in-depth [1]. He argues that it is appropriate to assume that in most social circumstances, privacy preferences are under constant change that means that they change with the context and the experience. The central assumption, which covers this perspective, is that people have on the one hand the desire to disclose personal information, for instance for the purpose of expecting informed advise or the pleasure of personal bonding. On the other hand, they have the desire to limit the disclosure of information to protect themselves from the judgment of others. Precisely, the balancing of the two poles is what is seen as a continuous adjustment process. Palen and Dourish refer to Altmans theory and say that privacy is a "boundary regulation process where people optimize their accessibility along a spectrum of openness and closedness depending on context" ([3], p. 130). Altman names different factors, which have impact on the desire to change how personal information is shared. One source of experience people gather during social interactions are, for instance, observations about how the own control decisions have impact on the systems behavior and how information about themselves are shared with others.

In the interaction with other humans, we continuously manage our boundaries by explicitly labeling information as sensitive as well as using social cues to indicate that certain information should be kept confidential. For instance by "changing intonation and speaking volume, or using posture or gestures" ([2], p. 29) as well as activities involving objects in the territory e.g. closing the doors. These implicit mechanisms are challenging to apply in a human-robot interaction because a robot, which is reacting appropriately to such mechanisms, requires a rich understanding of social norms.

Considering social norms (with respect to the handling of personal information as also referred to as informational norms [10] [11]) is in so far important for interface design, because norms can serve as indicators for the initial expectation users might have with respect to the handling of their data. Types of norms can be moral norms, conventions of etiquette or rules and procedures which allow a certain functioning in living together, for instance traffic rules (see: [10], p. 137). An example with pet-like robot companions could make clear how informational norms can be considered in the design. When you share an experience with your pet-like robot, which has personalization functionalities, it could be favorable if it could remember the experience and refer to it later. For instance you may have taught the robot, that a certain object is your "favorite yellow ball" (examples for robot object learning: [12], [13]). We imaging that you have a second robot (maybe of the same brand) you would most likely expect that the experience you have shared with one is not automatically referred to in interaction with the other. A robot-entity that appears as partly autonomous actor, similar to a pet, might be seen as something which keeps information to itself to a certain degree. It does not mean that a between-robot-exchange of shared experience is entirely inappropriate. However, respecting such an informational norm (which likely initializes user expectations) would mean to at least explicitly inform people of this kind of data handling and provide options to control it.

IV. Application: Practical Relevance

As an example for a robotic scenario in which we think considering the requirements of a dynamic boundary management is relevant, we refer to our work: We have setup a prototypical robot infrastructure, which is designed to support people working together in an office space.

The infrastructure contains two customized MetraLabs robots (Figure 1), which are able to move freely in the office space. Via keyword detection people can start an interaction with the robots. A face recognition functionality allows the robot to personally respond to the person who has started the conversation.

One example of an assistant the robot provides, which has the goal to support cooperation between people in their daily tasks, is a "looking for colleagues" functionality. It allows people to ask the robot where they can meet a certain person in the office. The robot keeps track of the encounters with persons in the office to be able to answer these questions. Having stored every encounter, the robot could response: "I have seen the person you are looking for in the kitchen 10 minutes ago" or give a general answer about the most likely location of that person based on accumulated observations.

This functionality is planned to be extended in a way that the robot provides additional information about the person the requester is asking for. For instance, information about topics, the requested colleague is currently working on or private information such as hobbies and preferences, which can facilitate social bonding. The information that is used based on past conversations with the robots. Furthermore, this will be extended with some intelligent matchmaking. This could facilitate the conversation when the two colleagues actually meet and thus support the initiation of a potential cooperation.

Regarding this example-application, we expect that the robot's behavior has impact on the social dynamics in the office where the flow of personal information plays a key role. We think a continuous management of the information flow between human and robotic system is important because it is very likely hard to predict for people before actually using it, which privacy impact a certain functionality has in run-time. This is especially the case when the functionality is based on *intelligent* processing capabilities.

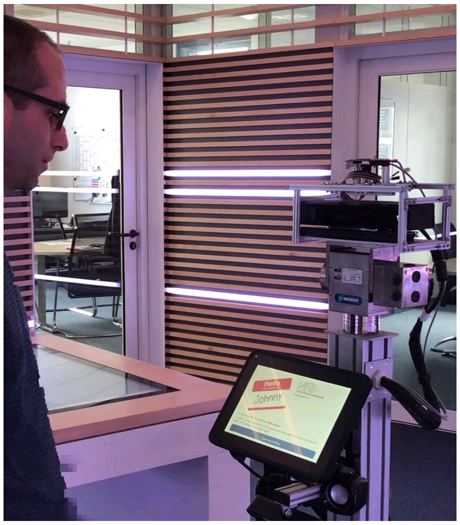

Fig. 1. A researcher in interaction with one of the office assistant robots called "Johnny".

In the next section, we discuss which methods in the design of social robotics technologies could be applied with respect to the possibilities of a dynamic boundary management.

V. Designing for Privacy: How to Achieve A Successful Boundary Management?

In this section, we will sketch how the requirements of a dynamic boundary management could be taken into account in a user interface design of a social robotic system.

Arguing for the need of a continuous management goes in line with authors criticizing technology designs which expect users to configure their data processing preferences (when the system is complex even in an excessive way) without knowing their in-situ needs. For instance, Lederer et al. stress, "setting explicit parameters and then requiring people to live by them simply does not work" ([4], p. 20). One reason they name is that it is a challenge for humans to "predict their [privacy] needs under hypothetical circumstances".

Having this in mind, the first major goal for a privacy-preserving design is to provide reasonable default settings and only few, easy to grasp control options. As mentioned in the theoretical section, informational norms determine users' initial expectation about how the system operates with respect to their data. Considering these norms in the interface design is a reasonable starting point to respect basic privacy needs but avoids overwhelming the user. A social robotic system, which by default shares personal data gathered with one robot with other robotic embodiments, does not fulfill the expectations most users would initially have. Consequently, such automatic exchange should rather not be the default setting or at least explicitly communicated to the user.

An easy to implement possibility to allow users to exercise control in a way, which takes account of the dynamics and context-dependence of privacy preferences are coarse-grained control options. Lederer et al. who discuss privacy management in a Ubiquitous Computing context call them "obvious, top-level mechanisms for halting and resuming disclosure" ([4], p. 22). An always-operable incognito button on a touch-screen attached to the robot (e.g. such a screen is shown in Figure 1) could do the job. Activating it could turns off all personal data processing functionality for a reasonable period of time.[3]

Beyond this coarse way of exercising control, more fine-grained configuration options could be useful to achieve a successful boundary management. In order to be able to reconfigure frequently, users have to be continuously aware of how data is processed in varying contexts for instance changing social constellations. Continuous awareness is a condition for exercising control in an informed or in other words meaningful way.[4] Regarding the application described before, users have a certain awareness of the information flow if they understand the basic functionalities of the robotic system: For instance the robot's ability to identify users via face recognition and the usage of the identity-data together with location information to allow other people in the office to access them via the "looking for person" function.

Continuing the example, even when people know about the described basic capabilities, they may not be aware of certain details, for instance that the robot is even identifying you when you just walk by and have not started an interaction. In the case of awareness gaps, the interface should be updated to give users feedback in order to fill the gap. In the example that could be realized by displaying a warning sign, for instance on a screen, which indicates that a tracking takes place. Ways to recognize users's awareness will be discussed next.

A. *Measuring Users' Awareness to Achieve Efficient Boundary Management*

Finding ways to gain knowledge about the users' awareness is valuable to design user interfaces, which allow them to perform an efficient boundary management. Next, we briefly

[3] For instance Gong et al. discuss a token which can be carried by users and serves as a "blackout button" in an Ubiquitous Computing application [14])

[4] For instance Florian Schaub stressed this relation as fundamental and highlights that "maintaining awareness" can be seen "a prerequisite for [meaningful] decision making" ([2], p. 35)

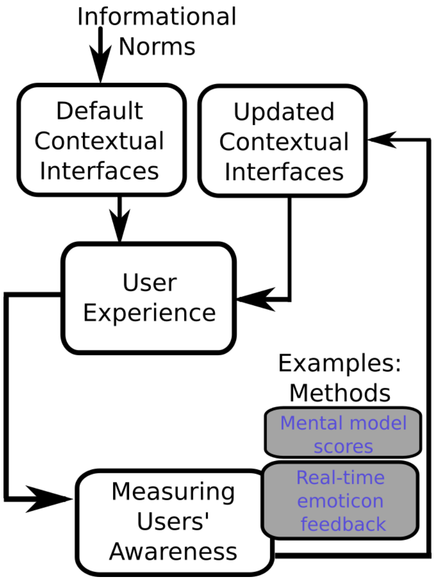
Informational Norms
Default Contextual Interfaces
Updated Contextual Interfaces
User Experience
Examples: Methods
Mental model scores
Real-time emoticon feedback
Measuring Users' Awareness

Fig. 2. A sketch of a design loop which shows how user experience could be evaluated in order to adapt/refine the human-robot interface to allow a successful boundary management.

describe which methods could be applied to measure users' awareness of the information flow and how that could be led back (see Figure 2).

One approach, we can apply to measure the awareness of the information flow is located in research on end-user interaction with machine learning applications. In this research it is investigated how explanations could help users to build adequate mental models of intelligent systems, for instance of recommendation systems or activity tracking devices [15] [16] [17] [18]. Kuluesza et al. evaluate the adequateness of mental models with respect to different types of explanations presented to the user about the internal mechanisms of the recommendation system [15]. Based on the participants' answers in the interviews and questionnaire the adequateness of mental models was evaluated with respect to a scoring metric. The metric is described as follows: "Participants' mental model "scores" were the number of correct minus the number of incorrect statements participants made during the experiment and on the post-study questionnaire, translated to a 0-to-10 (lowest-to-highest) scale" ([15], p. 6).

The individual's knowledge about the informational flow could be evaluated in a similar way. Questionnaires could be answered in-between interaction or afterwards. For the questions and comprehension tasks it could make sense to vary them around the parameters of an information flow.

Measuring user's awareness in an user experience study could help to refine the interface design in general. Imagining that the study results show that most people are not aware of when exactly an identification and location tracking takes place, signaling the tracking activity via an on-screen warning sign, as mentioned before, could increase their awareness.

An extension could be to measure users' awareness of the information flow not only by a user study rather utilizing awareness indicators directly in the run-time of the system in order to adjust (update) the feedback and control interface in an automatic and personalized fashion. The mentioned "mental model scores" were results of an extensive evaluation of participant's statements during the experiment. Adapting this idea by asking simple questions frequently in the run-time of the systems seems feasible to, less precisely, differentiate awareness from unawareness and directly update what is displayed on the interface.

Another idea is to provide users the possibility to communicate their *feelings* in a straightforward fashion with regards to the handling of personal data. This could be realized by allowing users to select an emoticon from an emotion feedback bar during interaction (Figure 3 illustrates how such a bar, e.g. displayed on the robot's touch-screen could look like). Based on users' emotional feedback, it seems reasonable to predict what could be their concerns. Consequently, the interface could be adjusted for instance by showing additional or less information, now and in the future, or highlight more prominently the control (configuration) options users have.

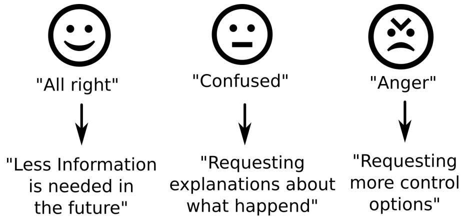
"All right"
"Confused"
"Anger"
"Less Information is needed in the future"
"Requesting explanations about what happend"
"Requesting more control options"

Fig. 3. Emoticons which could allow users to communicate their emotional states with regard to their agreement with as well as their understanding of the information flow. When people pressing one of the buttons in a certain situation, it can be predicted what user's emotional feedback could mean in the context. Understanding this relation could be used to adapt the individual's interface.

B. Limitations

Let us briefly mention one limitation of the proposed ideas to mediate users' boundary management. The requirements for a privacy-preserving design may be in conflict with the designers expectations about what looks appealing or which interaction modalities should be implemented. Explicit feedback about the information flow or possibilities for continues reconfiguration may disturb the immersion of human-like or pet-like companions. Considering this in more depth is a topic for future work. For instance work from Schaub at al. discusses different ways and modalities for giving feedback about personal data processing, also non-verbal and non-written, which could be beneficial to consider in this context [5].

References

[1] I. Altman, *The Environment and Social Behavior: Privacy, Personal Space, Territory, and Crowding*. Monterey, CA: Brooks Pub. Co., 1975.

[2] F. Schaub, R. Balebako, A. L. Durity, and L. F. Cranor, "A design space for effective privacy notices," in *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. Ottawa: USENIX Association, 2015, pp. 1–17.

[3] L. Palen and P. Dourish, “Unpacking ”privacy” for a networked world,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’03. New York, NY, USA: ACM, 2003, pp. 129–136.

[4] S. Lederer, J. I. Hong, A. K. Dey, and J. A. Landay, “Personal privacy through understanding and action: Five pitfalls for designers,” *Personal Ubiquitous Comput.*, vol. 8, no. 6, pp. 440–454, Nov. 2004.

[5] F. M. Schaub, “Dynamic privacy adaptation in ubiquitous computing,” Mar 2016.

[6] E. Union, “General data protection regulation,” *Official Journal of the European Union*, vol. L119, pp. 1–88, May 2016.

[7] S. Spiekermann and L. F. Cranor, “Engineering privacy,” *IEEE Transactions on Software Engineering*, vol. 35, no. 1, pp. 67–82, Jan 2009.

[8] M. Rueben, F. J. Bernieri, C. M. Grimm, and W. D. Smart, “Framing effects on privacy concerns about a home telepresence robot,” in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI ’17. New York, NY, USA: ACM, 2017, pp. 435–444.

[9] I. S. Rubinstein and N. Good, “Privacy by design: A counterfactual analysis of google and facebook privacy incidents,” *Berkeley Tech. LJ*, vol. 28, p. 1333, 2013.

[10] H. Nissenbaum, *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press, 2009.

[11] N. Nissenbaum, “A contextual approach to privacy online,” *Daedalus*, vol. 140, no. 4, pp. 32–48, 2011.

[12] P. Rouanet, P. Oudeyer, F. Danieau, and D. Filliat, “The impact of humanrobot interfaces on the learning of visual objects,” *IEEE Transactions on Robotics*, vol. 29, no. 2, pp. 525–541, April 2013.

[13] S. Hasler, J. Kreger, and U. Bauer-Wersing, *Interactive Incremental Online Learning of Objects Onboard of a Cooperative Autonomous Mobile Robot: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 1316, 2018, Proceedings, Part VII*, 01 2018, pp. 279–290.

[14] N.-W. Gong, M. Laibowitz, and J. A. Paradiso, “Dynamic privacy management in pervasive sensor networks,” in *Ambient Intelligence*, B. de Ruyter, R. Wichert, D. V. Keyson, P. Markopoulos, N. Streitz, M. Divitini, N. Georgantas, and A. Mana Gomez, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 96–106.

[15] T. Kulesza, S. Stumpf, M. Burnett, S. Yang, I. Kwan, and W. Wong, “Too much, too little, or just right? ways explanations impact end users’ mental models,” in *2013 IEEE Symposium on Visual Languages and Human Centric Computing*, Sept 2013, pp. 3–10.

[16] S. Stumpf, V. Rajaram, L. Li, W.-K. Wong, M. Burnett, T. Dietterich, E. Sullivan, and J. Herlocker, “Interacting meaningfully with machine learning systems: Three experiments,” *Int. J. Hum.-Comput. Stud.*, vol. 67, no. 8, pp. 639–662, Aug. 2009.

[17] B. Y. Lim, A. K. Dey, and D. 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*, ser. CHI ’09. New York, NY, USA: ACM, 2009, pp. 2119–2128.

[18] T. Kulesza, M. Burnett, W.-K. Wong, and S. Stumpf, “Principles of explanatory debugging to personalize interactive machine learning,” in *Proceedings of the 20th International Conference on Intelligent User Interfaces*, ser. IUI ’15. New York, NY, USA: ACM, 2015, pp. 126–137.