AI & SOCIETY (2022) 37:517–533
https://doi.org/10.1007/s00146-021-01210-3

ORIGINAL ARTICLE

Should my robot know what's best for me? Human–robot interaction between user experience and ethical design

Nora Fronemann[1] · **Kathrin Pollmann**[1] · **Wulf Loh**[2]

Received: 15 April 2020 / Accepted: 25 March 2021 / Published online: 28 April 2021

Abstract
To integrate social robots in real-life contexts, it is crucial that they are accepted by the users. Acceptance is not only related to the functionality of the robot but also strongly depends on how the user experiences the interaction. Established design principles from usability and user experience research can be applied to the realm of human–robot interaction, to design robot behavior for the comfort and well-being of the user. Focusing the design on these aspects alone, however, comes with certain ethical challenges, especially regarding the user's privacy and autonomy. Based on an example scenario of human–robot interaction in elder care, this paper discusses how established design principles can be used in social robotic design. It then juxtaposes these with ethical considerations such as privacy and user autonomy. Combining user experience and ethical perspectives, we propose adjustments to the original design principles and canvass our own design recommendations for a positive and ethically acceptable social human–robot interaction design. In doing so, we show that positive user experience and ethical design may be sometimes at odds, but can be reconciled in many cases, if designers are willing to adjust and amend time-tested design principles.

Keywords User experience · Human–robot interaction · Ethical design · Privacy · Autonomy · Elder care

1 Introduction

In the near future, assistive robots will become increasingly common in everyday life. The more robots are designed to share the same environment with humans, the more important it will be to consider the touchpoints between the two of them. To be fully integrated into the home and everyday environments, robots need to be equipped with capabilities to communicate and interact with us—preferably in a way that we can intuitively understand and react to. This is especially important in the context of elder care, which is often discussed as one of the main application areas for social robotics. Elderly users are more likely to face communication challenges such as impaired hearing and speech and tend to be more inexperienced with emerging technology. Therefore, they are often less likely to readily accept them in their daily lives or even their homes (Chien et al. 2019). Furthermore, to ensure the acceptance of robots, multiple aspects such as trust, expectations, and preconceptions towards them need to be considered when designing for elderly people (Frennert et al. 2020).

The interaction design should focus on the wellbeing and comfort of novice users as well as positive User Experience (UX). In the long run, these are the key factors for continuous use and acceptance of a social robot (Lamers and Verbeek 2011). For this, HRI researchers and designers have drawn upon existing usability and UX design principles that have been established for conventional software systems (Alenljung et al. 2017; Khan and Germak 2018). However, these design principles need some adjustment with respect to the physical presence and autonomous behavior of the robot. Unlike software systems, robots have a physical body. They are designed to move around and initiate interactions based on their own reasoning.

While there are a variety of interaction design and personalization strategies within the realm human–robot interactions (HRI) that can be used to enhance the UX and functionality of social robots, they come with certain ethical

✉ Nora Fronemann
nora.fronemann@iao.fraunhofer.de

[1] Fraunhofer Institute for Industrial Engineering IAO, Nobelstr. 12, 70569 Stuttgart, Germany

[2] Int. Center for Ethics in the Sciences and Humanities IZEW, University of Tuebingen, Wilhelmstr. 19, 72074 Tuebingen, Germany

repercussions. Especially two aspects need to be critically assessed: loss of control and the disclosure of personal data. A potential *loss of control* can manifest itself when the robot acts autonomously and the user is not able to fully understand or influence the course and outcome of the interaction (Sharkey and Sharkey 2012). Loss of control may be a serious ethical issue, opening the possibility for manipulation and paternalism (Darling 2017; Dworkin 2017). It can also lead to the infringement of fundamental ethical principles such as decisional autonomy, self-ownership, and one's own ideas about the good life (Frankfurt 1987; Turkle 2010).

On the other hand, not only personalized HRI applications that are based on a user profile require a fair amount of *personal data disclosure.* Already a robot's basic spatial orientation, movement, interaction initiation etc. are all dependent on constant sensory intake and aggregation. These in turn allow to draw many inferences on personal data about the user's home, daily schedule, lifestyle choices, and possibly health issues (Calo 2012; Rueben et al. 2017). A well-designed and seamless interaction might even exacerbate this issue by making it difficult for the user to realize that her privacy is at stake. In addition, manipulative interaction design can lead users to reveal more personal information than they are willing to. Privacy infractions are serious ethical issues, as they touch on the right to informational self-determination (Heesen 2017), as well as questions of maintaining one's personal identity (Roessler 2004).

In this paper, we show how, in HRI design practice, the two viewpoints of UX and ethics can inform each other and should thus be more closely connected. In doing so, we start from the practitioner's point of view to incorporate ethical decisions right into the developing process (Riek and Howard 2014). By balancing the user's well-being and comfort (as the fundamental UX design principles) with ethical principles such as autonomy and privacy, we explore the interdependencies between ethical and UX design and extend existing design principles to adjust both stances, and to deduce design recommendations for positive and ethical HRI design. Our considerations are based on use cases of social robotics in elder care developed within the NIKA project.[1]

We will start off in Sect. 2 with an overview of the research of social robotics and situate our own undertaking within this field. In Sect. 3, we describe the user-centered design principles that were used for the interaction design in the scenarios. These principles are well-established in usability and UX research, where they are frequently used to design for positive experiences with technology. With small adjustments, the same guidelines can be applied to social HRI, in order to design a robot's behavior from the user's perspective. Sect. 4 introduces two scenarios describing common interaction situations between a social robot and an older adult in a home setting. The scenario features Frank as a fictional generalized user ("persona") and NIKA as a fictional robot companion. We decided to use a fictional robot, in order not to be limited by the technical state of the art as well as the sensory and motoric limitations of actual robots. In Sect. 5, we discuss the described scenarios from an ethical perspective, with a focus on issues regarding loss of control and informational privacy. For both concepts, we highlight potential discrepancies between the initial—comfort and wellbeing-focused—interaction design and ethical principles such as autonomy, information privacy, and manipulation. Based on these discussions, in Sect. 6 we deduce design propositions to revise the initially proposed scenarios to balance UX design and ethical considerations. Finally, in Sect. 7 the main insights of our work are summarized as design recommendations for balancing ethical and UX design principles in HRI design.

2 Social robotics and elder care scenarios

Social robots are robots that operate in the vicinity of human beings and are designed to interact with humans. Cynthia Breazeal (2004) specifies the ability of social robots to engage in human-like communication as follows:

> "A social robot is able to communicate and interact with us, understand and even relate to us, in a personal way. It should be able to understand us and itself in social terms. We, in turn, should be able to understand it in the same social terms."

With the advance of social robots and their increasing application in daily life, it is crucial to carefully consider the future role of robots and the design of their interactive abilities that will allow them to fulfill these roles. Towards this goal, interaction design and ethics have to work hand in hand to promote robotic applications that match the users' expectations as well as their needs and are, at the same time, in line with societal and ethical considerations. On a fundamental level, these considerations include the question of emotional bonds with robotic systems and the questions of a good life (Sparrow et al. 2006; Turkle 2010; Seibt et al. 2016; Bryson 2018). Furthermore, designing social robots may result in emotional dependencies that could be easily exploited by companies and other actors (de Graaf 2016), or could have "serious harmful effects on the subject" when therapeutic or care robots are removed (Riek and Howard 2014).

Careful interaction design is especially important for novice user groups such as older adults who often lack the experience to be aware of these effects (Chien et al. 2019). At the

[1] Funded by the German Federal Ministry for Education and Research (BMBF),16SV7941, 16SV7944.

same time, there is a serious concern that these user groups are less likely to accept unfamiliar technologies, especially when an artificial social agent is suddenly introduced in a private space such as the home. It is thus important to also make sure that the robot's behavior is in line with the users' individual expectations and needs to ensure that it is accepted and used in the long term (Broadbent et al. 2009). Therefore, it can be beneficial to involve elderly people in the design process in order to understand their concerns, needs, and wants as well as to gather their feedback on design or interaction propositions (Fischer et al. 2020).

In the discussion on the design of robotic applications for older adults, applications for assisting actual care tasks enjoy a special focus (Coeckelbergh 2016). This paper, on the other hand, concentrates on applications that help to maintain elderly health and fitness and thereby preserve their self-reliance and autonomy as long as possible. Research has shown that this can be achieved by frequently engaging in leisure activity, especially if these activities are related to physical and cognitive training: A constant engagement in these kinds of activities has a positive effect on cognitive functions, overall well-being, and life quality (Everard 1999; Park et al. 2014). Physical and mental fitness can be increased (Sasidharan et al. 2006), and health-related problems such as diabetes, dementia, and cardiovascular diseases are reduced (Lotfi et al. 2018). While these activities may be beneficial from a health perspective, individuals are not always motivated to engage in them, especially over a long period of time. In this regard, carrying out such an activity together with other people has been found to positively affect motivation (Sasidharan et al. 2006).

Increasingly, however, elderly persons live alone and do not have the necessary social environment to do these exercises with a partner on a regular basis. In these cases, a social robot can be an alternative to support sustained engagement. A personal robot can encourage and motivate physical and mental exercises by acting as a sparring partner or coach, providing social interaction and companionship (Fasola and Mataric 2012). Being able to physically engage in activities together with the user, is a significant advantage that embodied technologies such as robots can offer over other technical devices (Lee et al. 2006).

To maximize user motivation and engagement, the robot needs to be able to adapt its behavioral strategies to the individual user's needs, abilities and preferences. This can be achieved by applying the principle of personalization (Dautenhahn, 2004; Syrdal et al, 2007). For this, many robots are designed as self-learning systems that can adjust their behavior to the interaction context and the user's individual needs and preferences. To this end, it is necessary to first develop a user profile—a model of the user that includes all relevant user characteristics that the robot needs to take into account (Karami, Sehaba and Encelle 2013). As a next step, different possible behavior patterns have to be designed for the robot. From these, the most suitable one is then chosen based on the user profile and context information (Pollmann 2019; Pollmann and Ziegler 2020).

3 Designing social human–robot interaction based on usability and user experience (UX) design principles

Designing behavior mechanisms for robots that foster a positive UX is challenging since all modalities like e.g. movement, gesture or voice need to be taken into account for an enjoyable and understandable interaction. Some attempts have been made to develop specific interaction principles for human–robot interaction. Drury (2004) proposed four design guidelines, namely "'enhance awareness' of the robots' immediate surroundings by providing spatial maps, 'lower cognitive load' by fusing sensor information automatically instead of letting the user fuse it mentally, 'increase efficiency' by providing a user interface that supports the control of multiple robots and 'provide help choosing robot modality' and level of autonomy by the operator". However, being based on computer-supported cooperative work, these guidelines provide insights into the technical specifications and control of the robot rather than its immediate interactive behavior. More recently, Cruz-Sandoval et al. (2018) proposed six interaction design principles: "human–robot interaction is reciprocal, robots interact in an enjoyable way, robots are machines, robots are unobtrusive, robots are inclusive and universal, robots do not create addictive behaviors, robots consider cultural, moral, and spiritual needs of users". While these principles capture important qualities of HRI, they are still very vague and do not provide any concrete suggestions of how to design robot behavior that provides positive interaction experiences.

In a series of studies, Kahn et al. (2008) deduced and evaluated eight appropriate behaviors for social robots in common, recurring interaction situations: "The initial introduction, didactic communication, in motion together, personal interests and history, recovering from mistakes, reciprocal turn-taking in game context, physical intimacy, and claiming unfair treatment or wrongful harms". While these design patterns are certainly contributing towards systematizing design goals and providing first instruction for HRI design, they are still very preliminary and can only be regarded as a first step. They also predominantly focus on verbal interaction between robot and user and have been established based on robot-child interaction scenarios.

In UX Design, the goal of designing intuitive, enjoyable interaction with technical systems is supported by various design guidelines and heuristics such as Nielsen and Molich (1990), Nielsen (1994), Shneiderman and Plaisant (2010),

Norman (2013). Although there is a rich literature on the subject, for reasons of brevity and consistency we mainly refer to their paradigmatic work. To create positive experiences while interacting with a robot, we adapted some of their well-known design guidelines and principles to the design of robot behavior.

Designing for the target group of elderly people, the focus on simplicity, comfort, and wellbeing, as well as the promotion of a seamless interaction is particularly relevant. This is especially true, since for this demographic a robot is a so far mostly unknown interaction partner. This design perspective is based on usability and UX standards, as well as different principles and heuristics. *Usability* as one of the main goals of good design is attained when a user in a specific context is able to achieve her specified goals with effectiveness, efficiency, and satisfaction (ISO 9241–11 2018). The second goal is to provide a positive *UX*, which is defined as a "person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service" (ISO 9241–210 2006). It includes the user's emotions, beliefs and perception—from the first confrontation with the product (even before use), until the product is discarded. Hassenzahl (2008) defines positive UX as a positive evaluative feeling when using the product. *We summarize the positive effects of these design principles on the interaction design under the term of "seamless interaction".*

The following paragraphs provide an overview of the Usability and UX design principles which are applied in the NIKA scenarios (cf. Sect. 4) and outline how they can contribute to HRI design.

The *efficiency of use* is an inherent part of the definition of "usability". Nonetheless, it is explicitly emphasized in many of the mentioned approaches, highlighting the fact that a system should be equally usable for any type of user, no expert knowledge required. Applied to intelligent systems such as robots, this might include self-initiated adjustments of the system to meet the user's individual abilities and characteristics.

To make a technical system easy and pleasant to use, interaction designers should also make sure to *minimize the user's cognitive load* (rather than increasing mental effort e.g. through complexity). This mainly means that the user's information intake and processing is supported by visualization and repetition of information. Thinking about assistive robots, this principle could even be extended towards smart robots that refrain from bothering the user with information that she does not need to complete a task. According to Druy et al. (2004), minimization of cognitive load could also be achieved by providing fused sensor information instead of making the user fuse the information mentally.

Consistency is another key requirement when designing any type of interactive system. Consistency demands that the system's actions and output are always expressed in the same way. In HRI design, specifically, it promotes the understanding of the system, making it comprehensible and predictable.

The design principle of *feedback* emphasizes the importance of always making the system status visible to the user. In social interaction with humans, we are experienced enough to automatically and effortlessly deduce what state a person is in or what she is going to do. However, this is usually not the case for technological artefacts, especially if we are confronted with a system we are not familiar with. Applying the principles of feedback to HRI means to make sure that the user receives appropriate feedback clarifying the internal information processing of the robot, especially in relation to user input. This requirement is necessary for the user to be able to freely engage in natural interaction with the robot.

Another important design principle is *error prevention*. This principle states that error-prone conditions should be eliminated, if possible, to avoid that the user runs into problems with the robot and may even have to do troubleshooting on her own. Robots are often cutting-edge technology that can be rather intimidating to less technology-accustomed users. Those tend to be hesitant regarding the long-term use of robots, among other things because they do not feel confident about the error handling.

The *conceptual model* refers to the user's mental model expressing her perception of how the system works based on the appearance and structure it offers. All other design principles and decisions feed into this mental model. As robots are complex systems, designers should be careful that the robot's features and interactive capabilities are in line with what its outer form and behavior suggests. If possible, a mental model that is similar to already internalized behavior patterns from the user's everyday environment should be facilitated. At the same time, these behavior patterns in combination with a certain humanoid outer form should not go as far as to raise expectations of, for example, sophisticated human behavior. For non-expert users, it is, therefore, advisable to design the robot in a way that the users can easily form a mental model that is not too complex. To do so, complex internal processes, such as the analysis of sensor data and algorithmic inferencing, should be hidden from the user.

The design recommendations to support the user-friendly robot design that we introduced so far mainly disregard ethical considerations. The proposed interaction design principles based on UX and Usability guidelines are applied in the next section to two companion-robot scenarios. In Sect. 5, we identify potentially critical points from an ethical perspective, and provide solutions to balance UX and ethical considerations in Sect. 6, which we note down as design recommendations in Sect. 7.

4 NIKA: a social companion robot for older adults

The scenarios depict scenes from everyday interaction in the context of companion robots in homes of elderly people, making use of a fictional persona called *Frank*, which is based on our empirical user analysis (NIKA 2019), and the conceptual companion robot *NIKA*. NIKA has been designed for elderly people still living in their own homes to prolong independent living. NIKA can remind its owner of appointments and routines, initiate and manage exercises like quiz games or physical training, motivate the person to be active and suggest activities. To do so, NIKA has a broad variety of sensors to detect the user and her state as well as the surroundings. NIKA is designed as a self-learning system: It can interpret the recorded sensor data and adapt its behavior to the situation and the user's preferences, in order to provide the interaction mechanisms and functionalities best suited for the individual user.

4.1 Scenario 1: getting to know each other

The fictional persona "Frank" is modelled after one of the five main target user groups that we deduced from qualitative interviews we conducted with elderly people living in a retirement community and in a multigeneration house (NIKA 2019). In addition, our insights from site-visits in a daycare for the elderly and a workshop with experts working in elder care are included. Frank is 82 years old and has been a widower for more than six years now. He has been getting along quite well, but lately, it seems to be more and more difficult for him to maintain the house and keep up with his daily chores. He is also getting a bit forgetful. Thus, Frank has agreed to his daughter Rachel's suggestion to buy the companion robot NIKA to keep him active and assist him with his daily routines. NIKA has now arrived at Frank's house and he unboxes it. Like with many current intelligent home devices, Frank receives an instruction manual including the terms of condition, which he can agree with by simply turning the robot on.

Although NIKA is a quite complex technological product and Frank has no experience with robots, the set-up process is made very *efficient* and *comfortable* for Frank. Through the voice interaction, Frank's *cognitive load is minimized*. He does not have to tackle and comprehend the internal processing of NIKA. His user profile is created, and filled in passing without any effort. Frank does not have to learn a complex interface, while NIKA automatically *prevents him from making errors*. The whole interaction is seamless and smooth. At the same time, NIKA provides *implicit feedback*. By starting to talk to Frank, NIKA does, for example, show that it is now active and starting to collect information about him. In summary, through the interaction NIKA promotes a very *simple conceptual model*, which is easy to comprehend for novice users like Frank, since it simulates a human interaction partner. This makes the whole unboxing and set-up process very pleasant and convenient for Frank.

4.2 Scenario 2: let's train your brain

NIKA knows from Frank's user profile that he would benefit from brain training, and that he likes to play games (Fig. 1a). NIKA has access to a database with a broad variety of entertaining games that also stimulate the brain. For each game, NIKA can act out different characters (such as co-player, opponent or coach), to motivate the user to perform well in the game and keep on playing. NIKA can track Frank's behavior through different sensors. Based on data analysis, NIKA can adjust the user profile over time and thus establish a more accurate profile for Frank. NIKA has, for example, learned from the data that Frank is not a morning person and usually not in the mood to play the quiz game before noon. NIKA has thus scheduled its suggestions for playing for the afternoon (Fig. 1b). After each interaction, NIKA records the relevant context configurations (time, Frank's activity level, Frank's mood) as well as motivational strategies that led to the desired (play the game) and undesired (rejection) reactions (Fig. 2g, h). Based on this data, Frank's user profile is updated, and new rules are created for the algorithm that initiates NIKA's interactions with Frank.

This scenario presents a similarly *seamless* interaction between Frank and NIKA, comparable to scenario 1 (Fig. 3). NIKA is driven by the mission of maintaining Frank's *long-term wellbeing* and therefore uses different motivational and interaction strategies (Fig. 2). As a self-learning system, NIKA automatically draws conclusions from Frank's responses and states, recorded by its sensors. Thus, the interaction is *efficient* and well-timed, tailored to Frank's individual needs and NIKA can avoid unwanted disturbances and prevent potentially unfavorable interactions. Frank's *cognitive load is minimized*, as he does not have to remember his daily exercise himself, nor does he have to make complex decisions about which method to use for brain training. Through *consistent behavior*, such as always using the same beep to indicate that it is approaching Frank, NIKA makes its behavior easily predictable for him.

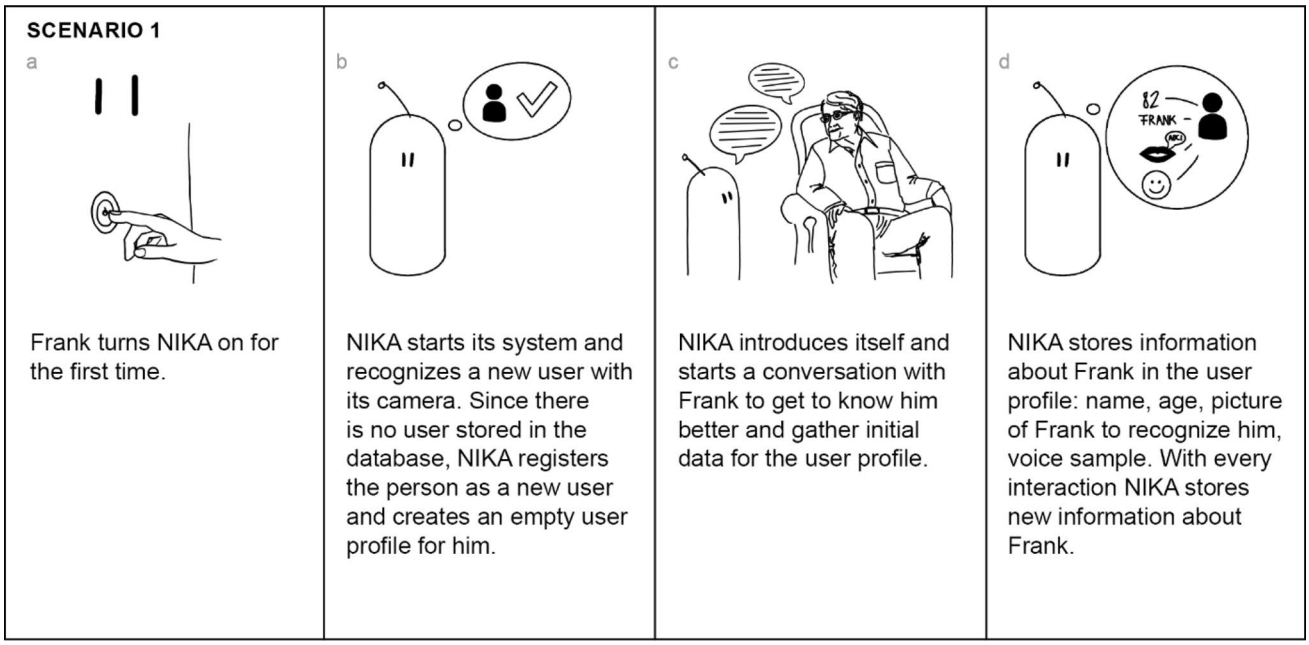

Fig. 1 Scenario 1: first encounter of Frank and NIKA

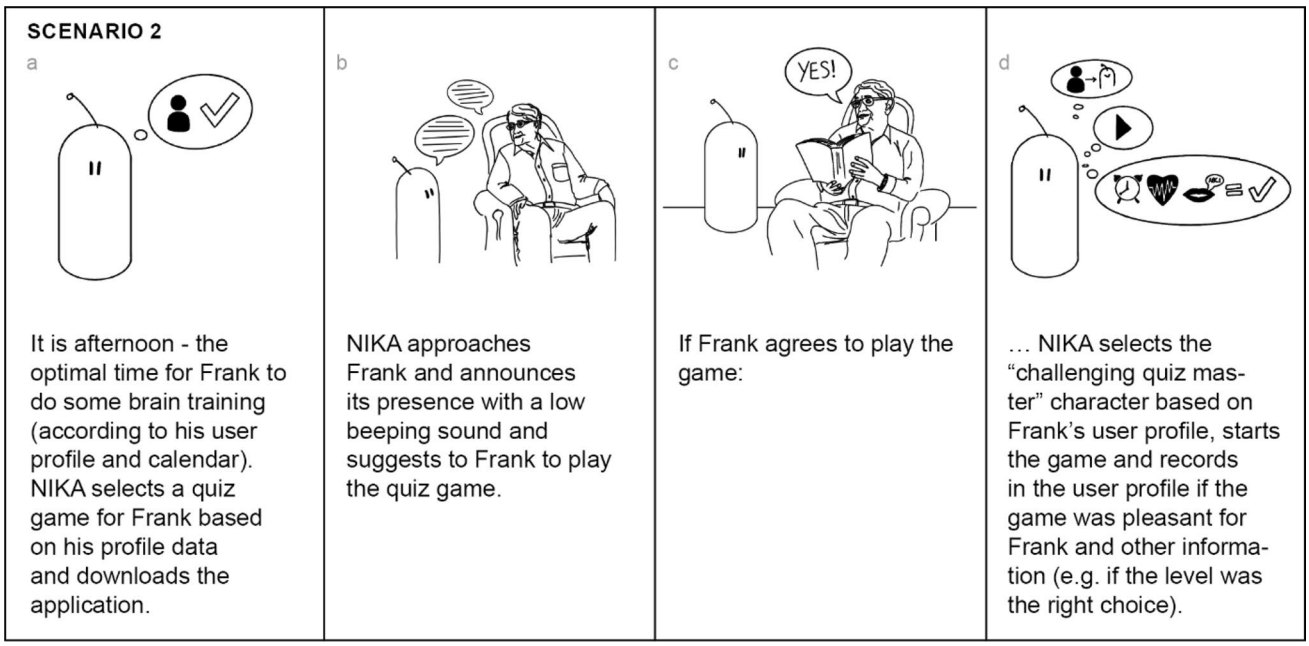

Fig. 2 NIKA wants to motivate Frank to play a quiz and Frank agrees

5 Ethical considerations for social HRI design

Based on the two scenarios, we have shown how *seamless interaction* and the underlying,well-researched and -tested Usability and UX design principles play out in day-to-day interactions. However, while the proposed interaction design of NIKA clearly contributes to Frank's comfort and well-being, these design decisions can conflict with ethical considerations (Coeckelbergh 2015). Concretely, seamless interaction may infringe upon the user's informational privacy and contribute to loss of control (Sparrow and Sparrow 2006; Calo 2012). For this reason, we have to take a second look at the scenarios from an ethical

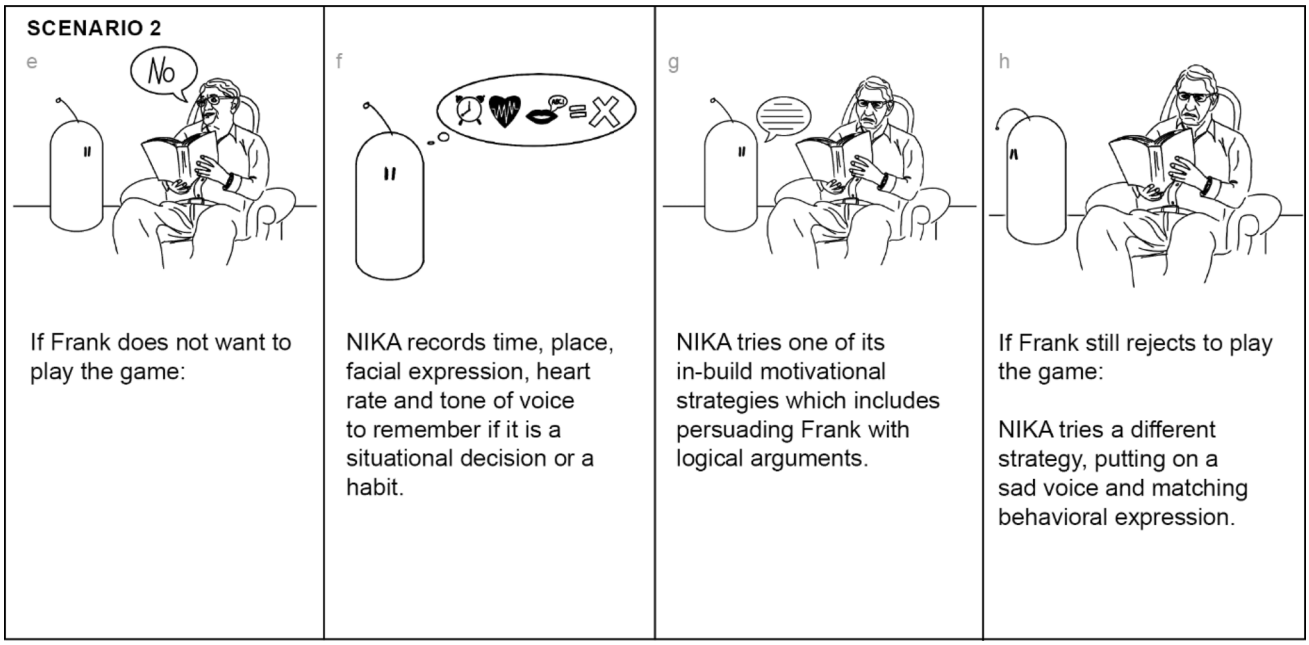

Fig. 3 Frank rejects to play a quiz, NIKA uses the motivational strategy of logical arguments and emotional expression

perspective, which might lead to a re-interpretation of the aforementioned design principles and—based on this—a redesign of the interaction design in the scenarios.

5.1 Informational privacy

Social robotics in all its instances has serious repercussions for informational privacy, for two reasons: First of all, robots are typically equipped with a variety of sensors to locate themselves in space and interact with their environment (Denning et al. 2009; Lutz et al. 2019). As these sensors are crucial for their basic functioning, they record data continuously, many times 24 h per day, effectively enhancing the possibility for "direct surveillance" (Calo 2012: 189). What makes this datafication especially problematic is the fact that social robots operate within close proximity of human users, often within their home. As a result, they potentially record a lot of sensitive data that is protected by special provisions and constitutional rights (inviolability of the home, the privacy of telecommunications and correspondence, right against self-incrimination, right to bodily integrity, in addition to the general fundamental right to privacy). In this regard, social robots increase the "access to historically protected spaces" (Calo 2012: 190).

Secondly, in order to realize interaction that is tailored to the individual user, as described in scenario two, the user has to disclose personal data that is then permanently stored, processed, and aggregated, such as her name, facial/voice recognition, preferences, and daily routines, acquaintances and friends, and the layout of her home. As these data sets become more comprehensive with usage time, they might produce medically relevant data series such as eating and sleeping habits, sports activities or medication. In addition, they increasingly become subject to data security issues and deanonymization techniques.

For these reasons, it is—from an ethical perspective—not sufficient to grant individuals the right to informational self-determination in the form of control over their own data. Although many philosophical theories of privacy emphasize the control element (Roessler 2004; Tavani 2007), particularly in the realm of social robotics the complexity and sheer amount of datafication is oftentimes very hard to monitor for the user. This is especially true for vulnerable groups such as the elderly, who often suffer from some form of cognitive or sensory impairment, while not being particularly technology-savvy.

In order to adequately evaluate the two scenarios from an ethical perspective, we will understand informational privacy in terms of an "appropriate flow of information" (Nissenbaum 2010: 149) within a certain context. Subsequently, the appropriateness of privacy norms varies with the respective context. For example, Frank's doctor may know his medical history but not his vacation plans, while Frank's boss is allowed to know his vacation plans but not his medical history. To preserve the "contextual integrity" (Nissenbaum 2010: 149) within each context, there has to be a common understanding of the privacy norms in place as well as the possibility for all affected—especially for the datafied subjects—to meaningfully express their interpretations of these norms.

In the scenarios, however, as with social robotics in general, there is yet only very basic societal agreement as to what constitutes an "appropriate flow of information", since social robotics is still in its infancy (van Nus 2016; Rueben et al. 2017). As long as this is the case, we argue that Frank should be empowered as much as possible to meaningfully exercise his data autonomy (Prainsack 2019; Hummel et al. 2020). While e.g. legal provisions such as the GDPR or national data protection laws may be silent about or even allow certain datafication practices, from an ethical point of view they may still be objectionable (Floridi 2016; Loh 2018; Seubert and Becker 2019).[2] To secure data autonomy, one of the ethical design guidelines could be to design the interactive behavior of the robot NIKA in a way to enable and facilitate his consent to the datafication, possibly through cascade consent models (Loe et al. 2015), privacy by design (Cavoukian 2011; O'Connor et al. 2017), and other technical means (systemic data protection). Especially in scenario 1, Frank's informational privacy is at issue by the fact that NIKA instantly starts collecting data and stores them in a user profile. Although legally, Frank may have consented to this datafication, he cannot meaningfully exercise his data autonomy if he does not fully understand what he actually consented to and what are the repercussions of this consent.

For this reason, NIKA's interaction with Frank should be designed in such a way that it *minimizes his cognitive load especially with respect to the various forms of datafication he is subjected to*. Only then will his approval of the datafication really amount to *informed consent*. This adjustment of the "cognitive load" design principle will in instances be in contrast with the overall goal of seamless interaction, as the minimization of cognitive load with respect to informational privacy will sometimes have to disrupt interaction, e.g. by switching to a meta interaction level.

5.2 Loss of control

The notion of "loss of control" (Sharkey and Sharkey 2012) serves as a common denominator between the ethical implications of infringements on the user's individual *autonomy* and the individual, user-centered perspective *feeling of alienation and loss of control*. In the proposed scenarios both aspects may be endangered by seamless interaction, for example when the robot adapts its behavior autonomously and the user is not able to predict and understand this adaptation. In these instances, the user is no longer in a position to effectively influence the course and outcome of the interaction.

From an ethical perspective, in such cases the user's individual autonomy is at stake (Kant 1785; Mill 1859; Sorell and Draper 2014), typically in the form of individual preferences and choices as well as notions of self-ownership (Korsgaard 1996; Susser et al. 2019). We assume that the robot is meant to support the user in her daily tasks and routines, and there are no egregious violations of design principles in place. In this case, we can expect possible limitations of the user's autonomy to be a result of the balancing in the behavior of the robot between subjective preferences and the user's perceived health and wellbeing (Beauchamp and Childress 1979; Nussbaum 2006; Yew 2020). In scenario two for example, NIKA uses mechanisms of self-learning and self-initiated adaptation to tailor its interaction and functionalities to Frank's individual preferences based on his user profile. The underlying algorithm is designed to balance Frank's needs and preferences with what is defined as Frank's desirable health and well-being. This was the goal to introduce the robot in Frank's home in the first place. To reach this goal, NIKA applies different motivational strategies that encourage Frank to continuously practice his brain training and thus maintain his good health.

Whether this balancing is in an ethically objectionable way *paternalistic*, depends on various factors (Fogg 2002). "Paternalism" consists in an "interference […] with another person, against their will, and defended or motivated by a claim that the person interfered with will be better off or protected from harm" (Dworkin 2017). This interference may occur either through manipulation or coercion.[3] In the context of social robotics, coercion is very rarely considered a viable design strategy (Santoni de Sio and van Wynsberghe 2016), and manipulative strategies are typically rather subtle. Still, even if the latter is the case and the interaction design and programming of the robot are well-intentioned and in the best interest of the user, her own alignment of goals, values, and intentions are in this case compromised (Frankfurt 1971).

[2] While law and ethics oftentimes overlap, this is not necessarily the case. Law not only has additional social coordination functions (such as e.g. traffic rules) that are not inherently ethical or unethical (Raz 1979). In addition, according to legal positivists, legal norms are justified through a different process than ethical norms (Hart 1961; Luhmann 2004) and therefore constitute a different kind of normativity than ethics (Habermas 1996; Korsgaard 1996). In other words, the "game of giving and taking of reasons" (Brandom 1998: 141) in ethics and law belongs to different "language games" (Wittgenstein 1953). In what follows, we are not concerned with the implications of individual data protection laws, but rather the ethical implications of "contextual integrity" with respect to informational privacy.

[3] While "coercion" in this context mainly refers to threatening or blackmailing the user into doing φ (Nozick 1969), by "manipulation" we mean the act of persuading the user to believe x or do φ, not by providing reasons or employing coercion, but through misinformation and/or exploiting psychological effects such as 'guilt trip', 'gaslighting', 'peer pressure', 'negging', or 'emotional blackmail' (Noggle 2018).

Especially problematic are cases of *strong paternalism.* Here, the coercion or manipulation of the user is well-intended, but towards goals that the user herself does not embrace. Rather, the manipulating agent tries to bring her to do or believe what "would be best for her", in the sense of what the manipulating agent thinks she has the most reason to do, all things considered (Parfit 1984; Williams 1981). This form of paternalism lies at the heart of many discussions about the possibility to favor the user's well-being at the expense of her autonomy (Halpern et al. 2007; Sharkey and Sharkey 2012; Sparrow and Sparrow 2006). Within human–machine interaction design, this typically occurs when the interaction designers too readily assume what would be best for the user without giving her the possibility to voice her own preferences.

With respect to the two NIKA scenarios above, these anticipated goals concern Frank's health or individual well-being, as it can be assumed that he bought the robot for these reasons. While the designers may assume that Frank *should* be healthy and fit—and therefore design NIKA in a way to try to manipulate him into leading a healthy life—they can never be certain that this is what Frank himself *wants* to do. In this sense, the emotional blackmailing in scenario 2 may constitute a form of strong paternalism, unless the programmers/designers make sure that NIKA acts according to Frank's explicitly voiced preferences (cf. below). In this case, NIKA uses design elements like voice and mimicry to mimic a heavily disappointed and sad person, to manipulate Frank into playing a quiz.

While according to some authors this form of paternalism is frequently employed by human caregivers in daily routines in elderly homes and day care (Gallagher 1998; Fernández-Ballesteros et al. 2019), from the perspective of ethical robot design any form of strong paternalism should be avoided (Sorell and Draper 2014). Artificially intelligent systems lack the proper judgment, empathy, and professional knowledge necessary to make such far-reaching decisions that deeply interfere with the user's autonomy (Häyry 1991). If ever, strong paternalism should be reserved for human interaction (Husak 1981; Misselhorn 2013; European Parliament 2017).

This changes with instances of *weak paternalism*, in which the robot manipulates or coerces according to reasons, interests, and preferences that the user has explicitly voiced. In this case, the user wants to act according to the same reasons that the designer has for her paternalizing action and suffers only from a lack of self-efficacy (Davidson 1970). With respect to the scenarios, this is e.g. the case when NIKA reminds Frank that he wanted to play a quiz once a day to keep himself mentally active (scenario 2, Fig. 1). Here, the database shows that Frank intends to keep mentally active and wants to do brain training, but has forgotten about it or just does not feel like it at the given time. Depending on how the personalization is realized, it may be safe to assume for the programmers/designers of NIKA that Frank has really formed this intention.

Whether forms of weak paternalism can be justified, highly depends on 1) how strong the incentivizing/coercive force is that the robot applies, and 2) how much its actions explicitly aligns with the preferences and interests of the user. Therefore, when NIKA initially persists in scenario 2 and tries to persuade Frank with logical arguments, it does so in alignment with Frank's own preferences, although it is clearly against Frank's explicit will at the moment. Since the incentivizing force is not very strong (persuasion), the interference can be justified from an ethical perspective. To fulfill the second condition from above (alignment), it is not enough that the intentions of the programmers align with Frank's, because the robot's behavior then may still appear to Frank as a black box (which also goes against the UX design principle of transparency). Therefore, the reasons for NIKA's insistence—i.e. reasons for the designers to implement a form of weak paternalism—should be made as clear to the user as possible. This could be done by giving feedback that Frank has previously mentioned that he likes to be reminded of playing a game or has previously liked playing the game.

In contrast, the emotional blackmail in scenario two potentially violates condition 1), i.e. the severity and persistence of the psychomotivational or emotional effect of the manipulation. Depending on how disappointed/ sad NIKA is made to act, the emotional blackmail may not only constitute a strong psychological effect. In addition, as the effect is hidden from Frank and consequently cannot be assessed by him as a motivational strategy, it is likely also deceptive (transparency requirement). On top of this, making Frank believe that NIKA is a human or an animal can also be "deceptive" (Grodzinsky et al. 2015) on a different design level, and should only be used under certain circumstances.[4] Therefore, even if condition 2) (alignment with Frank's own explicit preferences) is fulfilled, this form of weak paternalism should not be employed.

[4] While Grodzinsky et al. allow deceptive behavior in certain exceptions for "benign" purposes, it is exactly the scope of this discussion of paternalism to assess which conditions have to hold in order to ethically allow forms of deception or manipulation. It is our conviction that the severity of the manipulation has to be proportionate to the benefit for the well-being of the manipulated, which is not the case in this example. For an argument along similar lines cf. (Yew 2020:4).

6 Balancing UX design and ethical considerations

Designing a system which focuses on *seamless interaction* is, from a design perspective, a desirable goal to make the interaction simple and intuitive, especially for novice users like Frank. Nevertheless, as argued in Sect. 5, without the appropriate adjustments of the aforementioned design principles, the principle of data autonomy can be violated, leading to infringements on *informational privacy*. At the same time, while seamless interaction can contribute to the situational comfort and long-term wellbeing of the user, they also illustrate how it may lead to a *loss of control*. In the scenarios in Sect. 4, NIKA minimizes potential distractions and disturbances, reduces the interaction effort for Frank to the essential interaction steps, and determines the right time, place, and strategy for initiating interactions. On the downside, as we have seen, this may constitute forms of paternalism that are generally inadequate, or at least unsuitable for a robotic system. Taking into account the ethical considerations outlined above, we adjust the initially presented design principles as follows:

- The principle of *feedback* should be extended within reasonable limits to explicitly include information for the user about interaction steps, triggered actions, used data, automatically drawn conclusions, and options to change them. This will not only enhance the user's data autonomy and the possibility to give informed consent to datafication but also minimize the manipulation potential—and thereby paternalistic effects.
- *Minimizing the cognitive load* needs to be re-interpreted to also focus on informational privacy, rather than minimizing the cognitive effort for the user per se. This can be achieved e.g. by breaking the relevant information down into smaller chunks appropriate to the situation/interaction step. It can also be necessary to remind the user after a certain time of the choices she has made and give her the option to reconsent. As a general framework, the design principle of minimizing cognitive load should not be used as an excuse to hide as many of NIKA's decision-making processes from the user as possible.
- *Error prevention* should also include errors with regard to informational privacy, e.g. by employing mechanisms to inform the user about which information is captured, which sensors are used, or which choices were made and how they can be changed. Only if the user is made aware of the datafication through the interaction design, she can react to privacy-related errors.
- *Consistency* can be increased by using standardized interaction design mechanisms, which enable users to recognize interaction behaviors and sensors used for datafication. This not only enhances data autonomy and prevents paternalistic behavior patterns, but also effectively reduces the need for explicit feedback after a certain usage time.
- *Efficiency of use* should also be readjusted according to the ethical considerations mentioned in Sect. 5. To keep the interaction efficient while not infringing on the user's informational rights, complex datafication processes and their outcomes such as outlined in the general terms of conditions need to be broken down into understandable units. With regard to potential manipulations, "efficiency" should not simply be equated with the hiding of interaction decisions and decision-making processes, as already mentioned in the context of "cognitive load". Rather, appropriate feedback and information has to be provided in order to make the user aware of certain interaction goals and premises.

The adjustment of the design principles requires a revision of the initial scenarios presented in Sect. 4. At the same time, the revised scenario description serves as an illustration of how to apply the adjusted design principles in practical HRI design. We propose to make the following adjustments to the initial scenarios, taking into account the importance of balancing the seamless interaction design with the ethical principles of informational privacy and loss of control.

6.1 Revised scenario 1: getting to know each other

Figure 4 shows the adjusted *Scenario 1*. When designing for user autonomy and control, it is insufficient to only provide *implicit feedback* as presented in the initial scenarios (Borenstein et al. 2017; Yew 2020: 5). While this approach is very unobtrusive and straightforward for Frank as a novice user, it is questionable whether it also contributes to his autonomy and long-term wellbeing. Therefore, we propose to include additional interaction steps for the robot where possible, to provide *explicit feedback* for the user's action. If NIKA, as originally proposed, indicates that it is turned on by starting to talk to Frank, he might not be able to perceive and mentally process the effect of his own actions on the robot. A mechanism for explicit feedback on his action, such as a light that indicates that the robot's internal state has been set to operation mode "on", does not prolong the interaction, yet helps Frank to better understand the way NIKA works and predict its actions and intentions in the future (Fig. 4a).

In the interaction as described in Sect. 2, Frank's informational privacy is potentially compromised, as the activation of different sensors is not made transparent. While for some people it might be obvious that a robot needs to activate sensors such as cameras to be able to move and fully

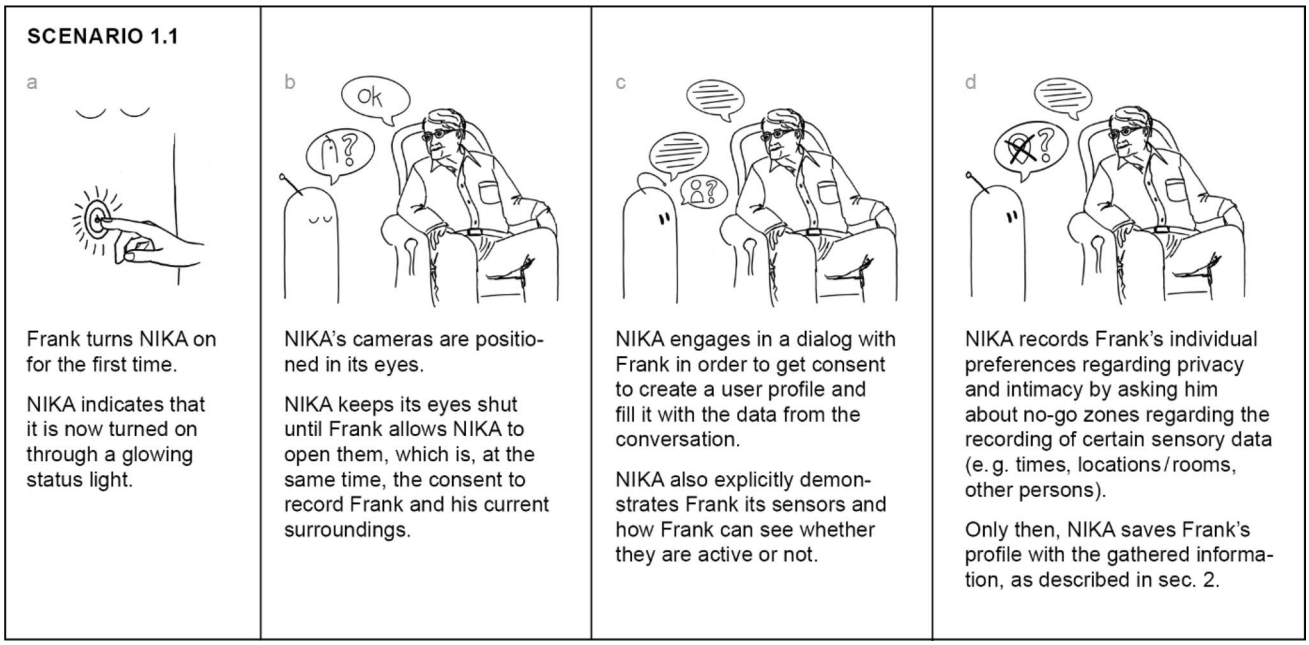

Fig. 4 Revised scenario1: Frank is using NIKA for the first time

operate, this cannot be taken for granted. Moreover, cameras integrated into robots hardly ever come with mechanisms to indicate whether they are currently actively recording or not. Even if Frank was aware of the constant and ubiquitous datafication, in the initial setup in Sect. 2 he would have had no say in it. In a private space such as the home, however, it is very likely that situations occur in which the user does not want the robot to record video. Small adjustments to the robot's design, such as putting movable eyelids in front of the camera hidden in NIKA's eyes (as e.g. the social robot "Miro" has them, but more for purposes of animal-like appearance than informational privacy and data autonomy), provide more transparency about the current data recording and plays into an intuitive *mental model* of the robot (Fig. 4b). Such mechanisms can be seamlessly integrated in the interaction flow, without negatively impacting the *efficiency of use* or *cognitive load*, and at the same time provide important information for the user, thus strengthening her data autonomy and informational privacy.

To enhance Frank's data autonomy and informational privacy by giving him the possibility to consent to the datafication in an informed way, these adjustments are only the first step to make the process of data recording more transparent. Frank also needs to understand how the recorded personal data is further processed during the long-term use of NIKA (Riek and Howard 2014). In the initial scenario, there is no mechanism to make transparent why, how and when NIKA's sensors are active, nor does the robot ask for Frank's explicit consent for the data recording and data storage in the user profile. Again, *explicit feedback* could be integrated into the interaction, to improve informational privacy and at the same time maintain an intuitive and straightforward interaction flow (Fig. 4c).

We suggest to slightly extend the introduction dialogue between NIKA and Frank, to acquaint him with the relevant interaction mechanism of the robot and give him the opportunity to make informed adjustments based on his need for privacy (Fig. 4d). While we propose that NIKA provides a verbal explanation about the creation of the user profile and data recording during the set-up session, the same information could be conveyed more implicitly in subsequent interaction situations. Thus, the *cognitive load* can be significantly reduced for future interactions. At the same time, NIKA *prevents* Frank from *making errors* in the future. This could be achieved by highlighting the sensors with light signals that are currently active and recording (the social robot "Pepper" provides e.g. the option to highlight ears and eyes, but so far this design feature is not used to demonstrate sensory recording). This is a very *efficient* way of gradually building the *conceptual model* of the robot's behavior (Fig. 4c). It is important to keep the interaction mechanisms *consistent* so that the user can link the explicit feedback in the beginning to the meaning of implicit feedback applied during longtime usage.

In the context of supporting informational privacy and reducing loss of control, the principle of *efficiency of use* can be adjusted according to this new perspective. The complexity of an intelligent system like a robot needs to be broken down into small comprehensive pieces to enable an efficient interaction. Privacy-related information that is

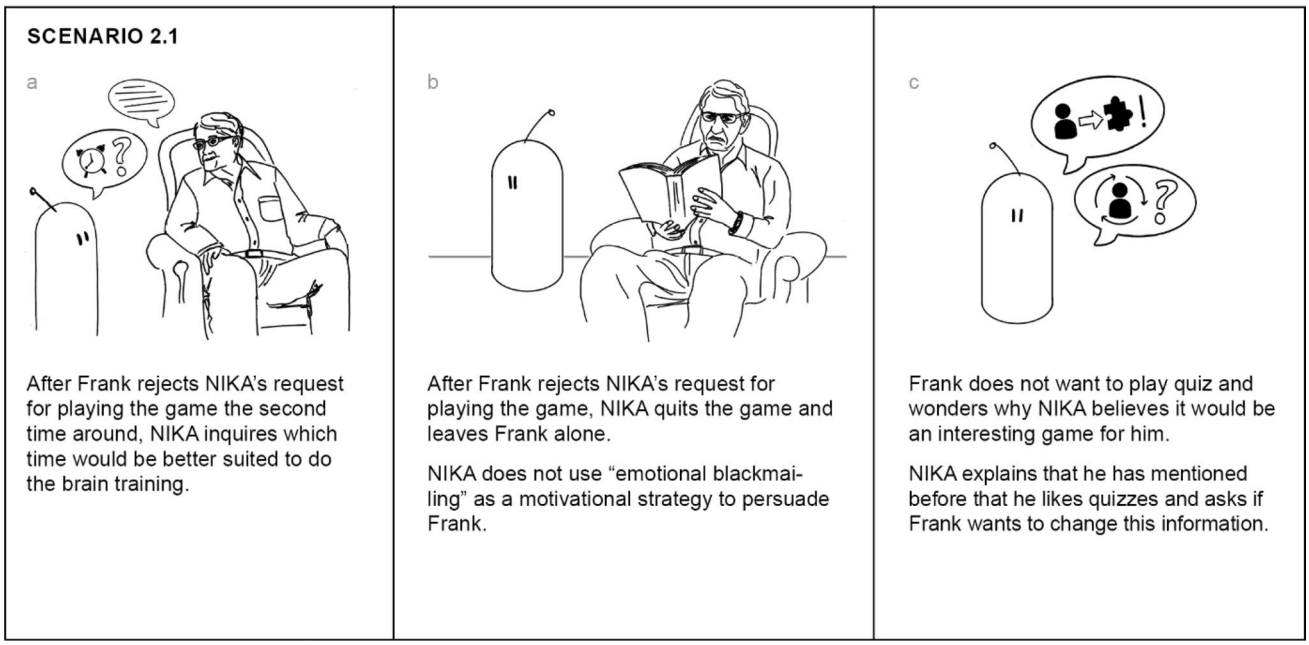

Fig. 5 Revised scenario 2: NIKA wants to motivate Frank to play a quiz

traditionally given by way of consent forms etc., could be designed as interactive dialogues between the user and the robot. Thereby, seamless interaction would be facilitated, which is at the same time informative as well as minimizing the *cognitive load*. Although it might seem that too much information during the set-up process can confuse Frank, there are two different goals with respect to cognitive load in conflict here: On the one hand, the goal of providing seamless interaction and enhancing error prevention, whereby Frank does not have to deal with a lot of different types of information that will excessively add to his cognitive load. On the other hand, by not presenting him with the information necessary to make an informed decision with respect to his datafication, NIKA raises his cognitive load in the sense that he has a hard time understanding what he is consenting to when presented with the general term of conditions or privacy statements.

To resolve this conflict, we suggest a *cascade model of informed consent*. NIKA guides Frank through the most important aspects step-by-step, thus making sure that he understands and knowingly agrees to NIKA's privacy policy and the GTC. This includes reminding Frank after a while of the choices he made during the setup, asking for his *reconsent*. In addition, we suggest *graded solutions* to many of the datafication issues. For example, many of NIKA's features will also be usable without collecting data from all sensors. Frank could use NIKA—albeit in a limited way—even if the cameras or microphones are shut off. By offering graded solutions, we circumvent all-in/all-out alternatives—which in reality do not amount to actual alternatives at all, since Frank's objecting to the datafication would only have the effect that he would not be able to use NIKA at all.

Furthermore, a separate app that visually rehashes Frank's privacy settings could be provided (as e.g. has been employed in common assistant technologies such as Siri,[5] Google Assistant[6] or Alexa[7]). Such an app not only presents the information through an additional medium (visual, instead of auditive in the case of the dialogue), it helps Frank to manage and keep track of NIKA's privacy settings. This additional way to make changes to them without having to enter into dialogic interaction with NIKA, supports his data autonomy, as the latter may be tedious (in cases of speech impediment) or inappropriate (when e.g. guests or therapeutic personnel is present).

6.2 Revised scenario 2: let's train your brain

In the second scenario, NIKA uses various motivational strategies to get Frank to play a quiz, some of which could conflict with ethical considerations. Thus, NIKA's behavior

[5] https://www.apple.com/de/siri/.

[6] https://assistant.google.com/.

[7] https://www.amazon.de/gp/browse.html?node=17084415031&ref_=nav_em__k_echo_privacy_0_2_7_17.

should be adjusted as shown in Fig. 5 (original scenario 2 Fig. 2).

When designing NIKA's behavior, we always have to consider that it was introduced with the goal to promote Frank's well-being and prolong his ability to live independently in his own home. We assume that in general Frank shares these goals. Therefore, even though some psychomotivational and emotional strategies that NIKA employs may be manipulative, they typically at most constitute a weak form of paternalism. This is because the goals of the manipulee and the manipulated align and the manipulee knows that they align (cf. Sect. 5). For this reason, the designers and programmers of NIKA can assume that if Frank is in a certain context not acting on those long-term goals, he is either not aware of them at this moment or may not have the necessary self-efficacy in a given context. In the described scenarios, the different *motivational strategies* that NIKA uses to persuade Frank to do his brain training could be therefore prima facie ethically acceptable. However, it might also be the case that Frank changed his long-term goals in the meantime; or that he has short-term goals that he deems more important at this moment than his long-term goals. For this reason, from an ethical perspective, we have to carefully examine the different strategies to evaluate how they might impact Frank's autonomy.

The strategy to convince Frank with arguments (Fig. 2g) is typically ethically sound, as long as the act of convincing (argumentative) does not subtly deteriorate into an act of persuading (emotional/ psychomotivational). While the former is autonomy-preserving, the latter may not be. On the other hand, emotionally blackmailing Frank (Fig. 2h) into doing his brain-training uses strong psychological effects as incentivizing force. Those are ethically problematic, since – as we have argued in Sect. 5 –to be considered an ethically acceptable weak form of paternalism, a psychomotivational or emotional effect may not cross a certain threshold of intensity and persistence. This holds even more since NIKA is an artificial robotic system and therefore lacks the proper judgment, empathy, and professional knowledge that a human caregiver typically has. In addition, such an interaction behavior also employs a form of "deception" (Grodzinksy et al. 2015), as the strong emotional response of the robot is likely to manipulate the user into believing that NIKA has real emotions (Turkle 2010; Bryson 2018).

For these reasons, we argue that the relevant threshold of intensity and persistence is crossed in Fig. 2h, hereby leaving the grounds of ethically acceptable weak paternalism. In order to avoid this, NIKA should not employ strong emotional or psychomotivational effects. Instead, it should leave Frank alone (Fig. 5b) and make a new motivation attempt at a later time (Fig. 5a). If the unwillingness to play the game persists, NIKA could address this and ask Frank whether he actually enjoys playing quiz (Fig. 5c). Even in severe cases of non-compliance, NIKA should rather inform a trusted third party than try to manipulate Frank using strong emotional or psychomotivational effects. While many emotional interaction cues often do not cross this intensity threshold, where this threshold lies heavily depends on the context and the "robot literacy" (Suto and Sakamoto 2014) of the user in question. Therefore, robot design should apply the precautionary principle here and program emotional behavior rather conservative, to avoid undue infringements on the user's autonomy.

With respect to NIKA's self-initiated choice of type of game and appropriate time for playing, this was introduced in Sect. 4 as a mechanism to reduce Frank's *cognitive load*. While this autonomous reasoning and actions of NIKA are very comfortable for Frank, we need to consider the consequences for his autonomy to prevent loss of control. As we have seen, deciding for Frank when best to schedule a quiz game is prima facie a good idea, but may very easily clash with Frank's short-term preferences at a given time. It is therefore crucial to make the robot's reasoning based on the user profile transparent to the user. Introducing NIKA's functionalities early in the set-up session will prevent misunderstandings and displeasure later on and thus increase the long-term *efficiency* and comfort of the interaction for Frank. To *prevent errors* made by a wrong interpretation of the gathered data about Frank, NIKA gives *feedback* on how the information can be adjusted (Fig. 5c).

7 Design recommendations and conclusion

In this paper, we demonstrated how well-established Usability and UX design principles can be applied to social HRI to design for comfort and wellbeing of the user. At the same time, we addressed how they might raise problems from an ethical perspective, especially regarding the privacy and autonomy of the user. We presented examples of how the original design principles and scenarios could be adjusted to balance UX design and ethical considerations in the interaction design of our companion robot NIKA. Based on the discussions and revised scenarios, we deduce five new design recommendations for a UX-driven and ethically sound social HRI design, which can support the realization of the adjusted design principles described above (Sect. 6). In what follows, we will shortly introduce these recommendations.

7.1 Transparency: make the internal state of the robot and the recording of sensory data as visible as possible and highlight sensors of the robot when they are actively recording

The current activation and internal processing state of the robot should be as transparent to the user as possible. This

starts with simple visualizations about whether the robot is turned on or off. In analogy with other technical devices, this can best be achieved with a status light, which is an unobtrusive way of indicating the robot's state independent of the use of other communication modalities. From a UX as well as an ethical point of view, it is important to provide transparency for the user with respect to the data that is recorded by the robot. In HRI, cameras and microphones are frequently used for user recognition and speech recording. Therefore, designers should make sure that it is visibly clear to the user when and what these sensors are recording.

7.2 Predictability: make sure the user knows what the robot is going to do next

The robot should display consistent behavior so that the user can predict how the robot is going to act. This enhances the user's autonomy as well as her feeling of control. Sounds and visualizations for implicit feedback should be easy to distinguish so that they can be internalized by the user without major effort. Predictability also facilitates error prevention, as the user can anticipate what the robot is going to do next. If the ears are illuminated and the eyelids open signaling that the voice recognition and the cameras are active, the user knows that her command to pause all recordings has not been processed by the robot.

7.3 Psychomotivational effects: make an informed decision about whether the robot should show emotionality

The emotionality of social robots is discussed controversially in HRI research (Hakli and Seibt 2017). While many studies suggest that an emotional robot might be perceived as more believable, interesting, and fun to interact with, emotionality can also be seen as deception and a mechanism of manipulation (as described in sects. 5 and 6). If designing emotional behavior for robots, designers have to think through the consequences in general; and particularly how emotionality might be used to manipulate the user into actions which diverge from her situational goals.

7.4 Step-by-step information: apply a cascading model of informed consent to enhance privacy

To be able to make informed decisions on the disclosure of data, the user needs to have suitable knowledge about what data is collected, as well as to understand the datafication goals and processes. A one-time all-in/all-out consent to datafication can, on the one hand, lead to cognitive overload, and is, on the other hand, ethically problematic, not least since the consent can only very rarely be considered informed. To remedy this, there are several ways, such as data minimization and privacy-by-design efforts (Cavoukian 2011; O'Connor et al. 2017). We recommend using a cascading model of consent, where the information is provided in smaller pieces and the user can give consent step-by-step (Loe et al. 2015). As seen in the revised scenario 1, this would mean that the robot negotiates the privacy settings as they appear in the interaction. Thus, the sensors of the robot are activated one after the other, also leaving the opportunity for Frank to deny the use of a certain sensor. By doing so, it minimizes cognitive load and maximizes control. Cascading informed consent in HRI relies heavily on designing alternatives to standard interaction patterns because prior to an agreement on a certain datafication framework, the robot should still be able to interact—albeit in a limited fashion. Moreover, if the user does not agree to a certain data collection, there should be other options to achieve the interaction goal.

7.5 Explainability: developing a mental model through feedback on sensors

To ensure that the user fully understands which sensors the robot uses, it is necessary to explicitly introduce each sensor and its function, asking the user for permission to turn it on. In addition to the *explicit* feedback, various forms of *implicit* feedback should be employed here. For example, the ears are illuminated while listening to the user. This combination of explicit and implicit feedback helps the user in the beginning to understand the functionality and usage of the sensors. After a while, the user intuitively recognizes the illuminated ear as the sign that the robot is listening. To safeguard informational privacy and maximize data autonomy during long-term usage, the robot should highlight the sensors again from time to time, asking for reconsent on their usage.

8 Future work

We are confident that the proposed adjustments and our design recommendations can play an important part in the puzzle to design social robots that address individual preferences and wellbeing, as well as comply with ethical considerations. To evaluate the proposed design principles, long-term studies are needed that assess how users perceive the interaction over a longer period of time. In our recommendations, there are still some uncertainties such as primarily relying on vocal dialogue to inform the user about privacy settings. These need to be evaluated. Especially interesting would be the question whether implicit feedback combined with reconsent are sufficient means to ensure the user's data autonomy and informational privacy. In this paper, we focused on the fictional persona "Frank" as the main user/buyer/benefactor of NIKA. Future research could

extend the discussion on balancing UX and ethical design towards interaction situations that include other actors, such as relatives, neighbors or visitors. Multi-user scenarios are also likely to occur in nursing homes or public spaces in general. A change in the social setting requires interaction designers to rethink their privacy design and consent models and come up with new behaviors of the robot that will also have to take into account the other person's goals, needs, and preferences. This might, for example, include a situational limitation of the robot's functionality, if some of the interacting persons did not agree to camera or voice recording. Multi-user scenarios also raise questions of who is controlling the robot, who can give the robot orders, and who has access to the recorded data. New research addressing these questions can build upon the adjusted design principles and proposed design recommendation of this paper.

Author contributions All authors contributed equally, read, and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. German Federal Ministry for Education and Research (BMBF), 16SV7941, 16SV7944–NIKA.

Availability of data and materials No data were used.

Code availability No software was developed.

Declarations

Conflict of interest Authors declares that they have no conflict of interest.

References

Alenljung B, Lindblom J, Andreasson R, Ziemke T (2017) User experience in social human-robot interaction. Int J Ambient Comput Intell 8:12–31. https://doi.org/10.4018/IJACI.2017040102

Beauchamp TL, Childress JF (1979) Principles of biomedical ethics. Oxford Univ. Press, New York

Borenstein J, Howard A, Wagner A (2017) Pediatric robotics and ethics: the robot is ready to see you now, but should it be trusted? In: Patrick Lin, Keith Abney, Ryan Jenkins (Eds.): Robot Ethics 2.0. New Challenges in Philosophy, Law, and Society: Oxford Univ Press, pp 127–141

Brandom RB (1998) Making it explicit. Reasoning, representing and discursive commitment, 2nd edn. Harvard Univ. Press, Cambridge

Breazeal CL (2004) Designing sociable robots. MIT Press, Cambridge

Broadbent E, Stafford R, MacDonald B (2009) Acceptance of healthcare robots for the older population: review and future directions. Int J of Soc Robot 1:319–330. https://doi.org/10.1007/s12369-009-0030-6

Bryson J (2018): Patiency is not a virtue. The design of intelligent systems and systems of ethics. In: Ethics and Information Technology 20 (1), S. 15–26

Calo R (2012) Robots and Privacy. In Patrick L, Keith A, George AB (Eds): Robot ethics. In: The ethical and social implications of robotics. Cambridge, Mass.: MIT Press, 187–202

Cavoukian, A (2011) Privacy by Design. The 7 Foundational Principles. http://scholar.google.de/scholar_url?url=https://www.ipc.on.ca/wp-content/uploads/Resources/pbd-implement-7found-principles.pdf&hl=de&sa=X&scisig=AAGBfm1NrsM8_Ee0B8ZoGrUt9bWYyg4Bcw&nossl=1&oi=scholarr. Accessed on 17 Oct, 2018

Chien S-E, Li Chu, Lee H-H, Yang C-C, Lin F-H, Yang P-L, Wang T-M, Yeh S-L (2019) Age difference in perceived ease of use, curiosity, and implicit negative attitude toward robots. J. Hum.-Robot Interact. 8:1–19. https://doi.org/https://doi.org/10.1145/3311788

European Parliament (2/16/2017): Civil law rules on robotics. EUP Resolution. 2015/2103(INL). Available online at https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html?redirect#title1, checked on 10/24/2020

Coeckelbergh M (2015) Artificial agents, good care, and modernity. Theor Med Bioeth 36(4):265–277

Coeckelbergh M (2016): Care robots and the future of ICT-mediated elderly care. A response to doom scenarios. In: AI & Society 31 (4), S. 455–462. https://doi.org/10.1007/s00146-015-0626-3 .

Cruz-Sandoval D, Favela J, Sandoval EB (2018) Strategies to facilitate the acceptance of a social robot by people with dementia. In: Kanda T, Ŝabanović S, Hoffman G, Tapus A (eds) strategies to facilitate the acceptance of a social robot by people with dementia. ACM Press, New York, pp 95–96

Darling K (2017): Who's Johnny? Anthropomorphic framing in human-robot interaction, integration, and policy. In: Patrick Lin, Keith Abney und Ryan Jenkins (Hg.): Robot Ethics 2.0. New Challenges in Philosophy, Law, and Society: Oxford Univ Press, S. 173–188

Dautenhahn K (2004) Robots we like to live with?!—A developmental perspective on a personalized, life-long robot companion. RO-MAN 2004: 13th IEEE international workshop on robot and human interactive communication. Piscataway, New Jersy, pp 17–22

Davidson, D (1970) How is weakness of the will possible? In: Donald Davidson: Essays on Actions and Events. Oxford: Clarendon Press 1980, 21–42.

de Graaf M (2016) An ethical evaluation of human–robot relationships. Int J of Soc Robotics 8:589–598. https://doi.org/10.1007/s12369-016-0368-5

Denning T, Matuszek C, Koscher K, Smith JR, Kohno T (2009) A spotlight on security and privacy risks with future household robots. In: Abdelsalam Helal (Ed.): Proceedings of the 11th International Conference on Ubiquitous Computing. New York, N.Y: Association for Computing Machinery, pp 105–114.

DIN EN ISO 9241–11:2018(en) Ergonomics of human-system interaction—Part 11: Usability: Definitions and concepts

DIN EN ISO 9241–110:2006(en) Ergonomics of human-system interaction—Part 110: Dialogue principles

Drury JL, Hestand D, Yanco HA, Scholtz J (2004) Design guidelines for improved human-robot interaction. In: CHI'04 extended abstracts on Human factors in computing systems, p 1540
Dworkin G (2017) Paternalism. Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/paternalism/. Accessed on 05 Jan, 2018
Everard KM (1999) The relationship between reasons for activity and older adult well-being. J Appl Gerontol 18:325–340
Fasola J, Mataric MJ (2012) Using socially assistive human–robot interaction to motivate physical exercise for older adults. Proc IEEE 100:2512–2526. https://doi.org/10.1109/JPROC.2012.2200539
Fernández-Ballesteros R, Sánchez-Izquierdo M, Olmos R, Huici C, Ribera Casado JM and Cruz Jentoft A (2019) Paternalism vs. autonomy: are they alternative types of formal care? In: Frontiers in Psychology 10, 1460.
Fischer B, Peine A, Östlund B (2020) The importance of user involvement: a systematic review of involving older users in technology design. Gerontologist 60:e513–e523. https://doi.org/10.1093/geront/gnz163
Floridi L (2016) On human dignity as a foundation for the right to privacy. In: Philosophy & Technology
Fogg BJ (2002) Persuasive technology. Using computers to change what we think and do. Elsevier, Amsterdam
Fosch-Villaronga E, Lutz C, Tamò-Larrieux A (2020) Gathering expert opinions for social robots' ethical, legal, and societal concerns: findings from four international workshops. Int J Soc Robot 12(2):441–458
Fracassi C, Magnuson WJ (2020): Data autonomy. Texas A&M University School of Law (Legal Studies Research Paper, 20–10).
Frankfurt H (1971) Freedom of the will and the concept of a person. J Philos 68(1):5–20
Frankfurt H (1987) Identification and wholeheartedness. In: Schoeman F (ed) Responsibility, character, and the emotions: new essays in moral psychology. Cambridge University Press, Cambridge, pp 27–45
Frennert S, Aminoff H, Östlund B (2020) Technological frames and care robots in eldercare. Int J of Soc Robot. https://doi.org/10.1007/s12369-020-00641-0
Gallagher SM (1998) Paternalism in healthcare decision making. Ostomy Wound Manag 44:24–25
Grodzinsky FS, Miller KW, Wolf MJ (2015) Developing automated deceptions and the impact on trust. Philos Technol 28(1):91–105
Habermas J (1996) Between facts and norms. In: Contributions to a discourse theory of law and democracy. MIT Press, Cambridge
Hakli R, Seibt J (2017) Sociality and normativity for robots. Springer, Cham
Halpern SD, Ubel PA, Asch DA (2007) Harnessing the power of default options to improve health care. N Engl J Med 357(13):1340–1344
Hart HLA (1961) The concept of law. Clarendon Press, Oxford
Hassenzahl M (2008) User experience (UX): towards an experiential perspective on product quality. In: Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine. ACM, New York, NY, USA, pp 11–15
Häyry H (1991) The limits of medical paternalism. Routledge, London, New York
Heesen J (2017) Informationelle Selbstbestimmung. Grundbegriffe der Kommunikations- und Medienethik (Teil 10). In Communicatio Socialis 50 (4), pp 495–500.
Hummel P, Braun M, Dabrock, P (2020) Own Data? Ethical Reflections on Data Ownership. In: Philosophy & Technology, pp 1–28
Husak D (1981) Paternalism and autonomy. Philos Public Aff 10(1):27–46
Kahn PH, Freier NG, Kanda T, Ishiguro H, Ruckert JH, Severson RL, Kane SK (2008) Design patterns for sociality in human-robot interaction. In: Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction. ACM, pp 97–104
Kant I (1785) Grounding for the Metaphysics of Morals. In: I. Kant , Ethical Philosophy, James W. Ellington, trans., Indianapolis, IA: Hackett Publishing Co. 1983
Karami AB, Sehaba K, Encelle B (2013) Adaptive and Personalised Robots—Learning from Users' Feedback. In: 2013 IEEE 25th International Conference on Tools with Artificial Intelligence. IEEE, 626–632. https://doi.org/10.1109/ICTAI.2013.98 .
Khan S, Germak C (2018) Reframing HRI design opportunities for social robots: lessons learnt from a service robotics case study approach using UX for HRI. Future Internet 10:101. https://doi.org/10.3390/fi10100101
Korsgaard C (1996) The sources of normativity. Cambridge University Press, New York
Lamers MH, Verbeek FJ (ed) (2011) Human-Robot Personal Relationships. Third International Conference, HRPR 2010, Leiden, The Netherlands, June 23–24, 2010, Revised Selected Papers. Springer Berlin Heidelberg, Berlin, Heidelberg
Lee KM, Jung Y, Kim J, Kim SR (2006) Are physically embodied social agents better than disembodied social agents?: the effects of physical embodiment, tactile interaction, and people's loneliness in human-robot interaction. Int J Hum Comput Stud 64:962–973. https://doi.org/10.1016/j.ijhcs.2006.05.002
Loe J, Robertson CT, Winkelman DA (2015) Cascading consent for research on biobank specimens. Am J Bioeth 15(9):68–70
Loh W (2018) A practice-theoretical account of privacy. Ethics Inf Technol 20(4):233–247
Lotfi A, Langensiepen C, Yahaya SW (2018) Socially assistive robotics: robot exercise trainer for older adults. Technologies 6:32
Luhmann, N (2004): Law as a social system. Klaus AZ, Fatima K (eds). Oxford Univ. Press, Oxford
Lutz C, Schöttler M, Hoffmann CP (2019) The privacy implications of social robots: scoping review and expert interviews. Mobile Media Commun 7(3):412–434
Mill JS (1859) On Liberty. New York: Norton 1975
Misselhorn C (2013) Robots as Moral Agents. In Roevekamp F (ed): Roboethics. Proceedings of the Annual Conference on Ethics of the German Association for Social Science Research on Japan. München: Iudicum, pp. 30–42.
Nielsen J (ed) (1994) Usability inspection methods. Wiley, New York
Nielsen J, Molich R (1990) Heuristic evaluation of user interfaces. In: Chew JC, Whiteside J (eds) Proceedings of the SIGCHI conference on Human factors in computing systems Empowering people - CHI '90. ACM Press, New York, New York, USA, pp 249–256
Nissenbaum H (2010) Privacy in context. Technology, policy, and the integrity of social life. Stanford CA: Stanford Law Books
Noggle R (2018) The Ethics of Manipulation. Edited by Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/ethics-manipulation/, checked on 6/6/2019
Norman DA (2013) The design of everyday things, Revised and expanded edition. Basic Books, A Member of the Perseus Books Group, New York
Nozick R (1969) Coercion. In: Sidney Morgenbesser/Patrick Suppes/Morton White (eds.): Philosophy, Science, and Method: Essays in Honor of Ernest Nagel. New York: St. Martin's Press, 440–472
Nussbaum M (2006) Frontiers of justice. Belknap Press of Harvard University Press, Cambridge
O'Connor Y, Rowan W, Lynch L, Heavin C (2017) Privacy by design. Informed consent and internet of things for smart health. Procedia Comput Sci 113:653–658
Parfit D (1984) Reasons and persons. Clarendon, Oxford
Park DC, Lodi-Smith J, Drew L, Haber S, Hebrank A, Bischof GN, Aamodt W (2014) The impact of sustained engagement on

cognitive function in older adults: the synapse project. Psychol Sci 25:103–112
Pollmann K (2019) Behavioral Design Patterns for Social, Assistive Robots-Insights from the NIKA Research Project. In: Mensch und Computer Workshopband. Gesellschaft für Informatik eV
Pollmann K, Ziegler D (2020) Personal Quizmaster: A Pattern Approach to Personalized Interaction Experiences with the MiRo Robot. In: Preim B, Nürnberger A, Hansen C (eds) Proceedings of the Conference on Mensch und Computer. Association for Computing Machinery, New York, NY, USA, pp 485–489
Prainsack B (2019) Logged out: Ownership, exclusion and public value in the digital data and information commons. Big Data Soc 6(1):205395171982977
Raz J (1979) The authority of law. Clarendon, Oxford
Riek L, Howard (2014) A code of ethics for the human-robot interaction profession. In: Proceedings of We Robot, S. 1–10
Roessler B (2004) The value of privacy. Polity Press, Cambridge
Rueben M, Grimm C, Bernieri F, Smart W (2017): A taxonomy of privacy constructs for privacy-sensitive robotics. ArXiv. Online verfügbar unter https://arxiv.org/pdf/1701.00841.pdf. Accessed on 05 Aug, 2020
Santoni de SF van Wynsberghe A (2016): When should we use care robots? The nature-of-activities approach. In Science and engineering ethics 22(6):1745–1760
Sasidharan V, Payne L, Orsega-Smith E, Godbey G (2006) Older adults' physical activity participation and perceptions of wellbeing: examining the role of social support for leisure. Manag Leis 11:164–185
Seibt J, Nørskov M, Schack A, Søren (Hg.) (2016) What social robots can and should do. Proceedings of Robophilosophy 2016/TRANSOR 2016
Seubert S, Becker C (2019) The culture industry revisited: sociophilosophical reflections on 'privacy' in the digital age. Philos Soc Crit 45(8):930–947
Sharkey A, Sharkey N (2012) Granny and the robots: ethical issues in robot care for the elderly. Ethics Inf Technol 14(1):27–40
Shneiderman B, Plaisant C (2010) Designing the user interface: strategies for effective human-computer interaction, 5th edn. Addison-Wesley, Boston
Sorell T, Draper H (2014) Robot carers, ethics, and older people. Ethics Inf Technol 16(3):183–195
Sparrow R, Sparrow L (2006) In the hands of machines? The future of aged care. Mind Mach 16(2):141–161
Susser D, Rössler B, Nissenbaum H (2019) Technology, autonomy, and manipulation. Internet Policy Rev 8(2):1–22
Suto H, Sakamoto M, (2014): Developing an education material for robot Literacy. In Sakae Yamamoto (Ed.): Human Interface and the Management of Information. Information and Knowledge in Applications and Services. Cham. Cham: Springer, pp. 99–108
Syrdal DS, Koay KL, Walters ML, Dautenhahn K (2007) A personalized robot companion?-The role of individual differences on spatial preferences in HRI scenarios. In: RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication. IEEE, pp 1143–1148
Tavani H (2007) Philosophical theories of privacy: Implications for an adequate online privacy policy. Metaphilosophy 38(1):1–22
Turkle S (2010) In good company? On the threshold of robotic companions. In: Yorick Wilks (Hg.): Close engagements with artificial companions. Key social, psychological, ethical and design issues. Philadelphia, PA: John Benjamins (Natural language processing, 8), S. 3–10
van Nus M (2016) Social robots, privacy, and ownership of data: some problems and suggestions. In Johanna Seibt, Marco Nørskov, Søren Schack Andersen (Eds.): What social robots can and should do. Proceedings of Robophilosophy 2016/TRANSOR 2016, vol. 290. Amsterdam: IOS Press (Frontiers in artificial intelligence and applications, 290), pp. 190–191
Williams B (1981) Internal and external reasons. In Bernard Williams: Moral luck. Philosophical papers 1973–1980. Cambridge UK: Cambridge University Press, pp 101–113
Wittgenstein L (1953): Philosophical investigations. With the assistance of Elizabeth Anscombe. Oxford: Blackwell
Yew GCK (2020) Trust in and Ethical Design of Carebots: The Case for Ethics of Care. International Journal of Social Robotics:1–17. https://doi.org/10.1007/s12369-020-00653-w

References from the NIKA Project

NIKA, 2019: https://www.nika-robot.de/zielgruppen/

Robots and Assistant Technologies

Softbank Robotics, Pepper: https://www.softbankrobotics.com/emea/en/pepper
Consequential Robotics, Miro-E: https://www.miro-e.com/robot
iRobot: https://www.irobot.de/roomba
Apple Siri: https://www.apple.com/de/siri/
Google Assistant: https://assistant.google.com/
Amazon Alexa: https://www.amazon.de/gp/browse.html?node=17084415031&ref_=nav_em__k_echo_privacy_0_2_7_17