Does a Loss of Social Credibility Impact Robot Safety?

Balancing Social and Safety Behaviours of Assistive Robots

Catherine Menon

Adaptive Systems Research Group
School of Computer Science
University of Hertfordshire
Hatfield AL10 9AB, United Kingdom
Email: c.menon@herts.ac.uk

Patrick Holthaus

Adaptive Systems Research Group
School of Computer Science
University of Hertfordshire
Hatfield AL10 9AB, United Kingdom
Email: p.holthaus@herts.ac.uk

***Abstract*—This position paper discusses the safety-related functions performed by assistive robots and explores the relationship between trust and effective safety risk mitigation. We identify a measure of the robot's social effectiveness, termed social credibility, and present a discussion of how social credibility may be gained and lost. This paper's contribution is the identification of a link between social credibility and safety-related performance. Accordingly, we draw on analyses of existing systems to demonstrate how an assistive robot's safety-critical functionality can be impaired by a loss of social credibility. In addition, we present a discussion of some of the consequences of prioritising either safety-related functionality or social engagement. We propose the identification of a mixed-criticality scheduling algorithm in order to maximise both safety-related performance and social engagement.**

Keywords–Human-Robot interaction; Social credibility; Robot safety.

I. INTRODUCTION

Assistive robots offer significant benefits to an increasingly elderly population, both in terms of their social impact and their functionality [1][2]. Assistive robots support independent living by aiding humans to conduct basic activities, such as preparing food and bathing. Similarly, these robots may support the psychological health of elderly or isolated individuals via socially-important behaviours, providing companionship and encouraging these individuals to engage and interact.

There are safety implications to the use of assistive robots, both in terms of the physical hazards they present and in terms of the functionality they provide. An assistive robot will often act as mitigation for a safety risk, alerting the user to a hazardous situation and requesting that they take action.

In this paper we bring together concerns from the safety community and the robotics community. The social effects of autonomous systems are not typically factored into hazard analysis of these systems, and this paper aims to address that omission. Equally, from an Human-Robot Interaction (HRI) perspective the ways in which the social performance of an assistive robot are affected by safety features (e.g., automatic stops, avoidance of physical contact) is not always explicitly considered. Bringing these concerns together within a single domain provides the research community with a foundation for discussing how to assure the safety of an autonomous system which must also perform another (social) function. This is relevant not only to assistive robots but also to autonomous vehicles, medical devices and companion robots.

Figure 1. The assistive robot Care-O-Bot 4 configured with two arms and spherical hip and head joints.

To meet this aim we examine how both the safety-critical and socially important behaviours of an assitive robot rely on the user's engagement with the robot. User engagement, particularly in safety-critical situations, is partially determined by the *social credibility* of the robot, or how well it follows social norms relevant to its environment. In Section II we present a case study assistive robot, identifying some of its socially important behaviours. Section III looks at the functional and physical hazards associated with such a robot, while Section IV considers restrictions on the behaviours considered socially appropriate, as well as introducing and defining the concept of social credibility. In Section V we identify how a loss of social credibility impacts both safety-critical and socially-important types of behaviour, illustrate how such behaviours may be in conflict with each other and discuss a solution which allows both to be prioritised. Section VI contains a proposal for future work to validate these concepts and solution, summarizes our position and concludes our contribution.

 ISBN: 978-1-61208-698-9

II. The Care-O-Bot assistive robot

The Care-O-Bot [3] is an up-to-date example of a mobile assistant robot with the capacity for social interaction. Its most recent iteration can be adapted to various applications in care due to its modular design. When equipped with two 7-DoF arms and two spherical joints at its hip and head (as shown in Figure 1), it can manipulate objects within an exceptionally large workspace. When such a robot operates within a sensorized domestic environment [4], it is able to support humans in their daily activities. In conjunction with its interactive capabilities the robot therefore is well suited to execute a wide range of desirable tasks in elderly care [5].

In such a setting, the Care-O-Bot might be typically expected to perform a range of functions including:

- Accepting and handling a parcel at the front door
- Reminding a user to take their medication
- Assisting a user to carry food items from the kitchen

In addition, more complex temporal behaviours [6] can also be defined by a formal or informal care-giver. These behaviours may include requesting the robot to alert a care-giver if the user has remained in bed for longer than a specified time, or alerting a user if the oven has remained on after cooking a meal. Existing research has utilised formal verification [7] to ensure that user-defined behaviours do not conflict with each other, and has highlighted a need for human-intelligible output to help users define behaviours. In addition to these care-giving behaviours, the Care-O-Bot would typically be expected to encourage the user to engage and interact by offering entertainment and companionship.

III. Safety critical performance of assistive robots

Some of the functions performed by an assistive robot such as the Care-O-Bot have the potential to impact safety. The robot presents both physical hazards (e.g., its weight can contribute to crush injuries) as well as functional hazards. Functional hazards are those resulting from its behaviour: the robot may fail to perform a safety-critical function (e.g., reminding a user to take medication) or may perform this function incorrectly (e.g., reminding the user too frequently).

The Care-O-Bot has been designed with safety as a priority. All personal care and assistive robots are required to comply with safety standards [8], as well as broader UK safety legislation [9]. The Care-O-Bot accordingly contains a number of features to reduce or eliminate collisions with a user [10]. The robot's base is equipped with three laser range sensors with a safety Programmable Logic Controller (PLC) that allow for a 360 degree obstacle recognition at ankle height. Its joints are protected with two separate safe-torque-off (STO) switches at base and torso. The STOs are either triggered by the laser range sensors, one of two emergency buttons at the robot's front and back (Figure 1), or a wireless emergency stop. Furthermore, the robot's autonomous navigation software implements well established collision avoidance mechanisms [11] by default. Despite this, however, there is a lack of sensors at the arm joint and thus no mitigation against crush injuries received at this site. As a result, Care-o-bot requires constant monitoring while participating in interactions with humans that involve the robot's arms.

A. System failure and resultant hazards

System failure still remains an issue for the Care-O-Bot, as for all safety-critical systems. Should the proximity sensors fail, the Care-O-Bot could collide with a user and cause injury. Other potential hazards include hot surfaces from the engine, trip hazards from the wheels, potential corrosive substances and the presence of electrical items. Furthermore, collision hazards are not limited only to collision with the robot itself, but include collisions with any objects it is holding. In particular, a key characteristic of the Care-O-Bot is the presence of arms that can be used to carry hot liquids on a tray [6]. Should a system failure occur, the arm may be stuck in an unpredictable position, resulting in anything held being spilt on the floor or on a user. It is clear, therefore, that complete or partial system failure of the Care-O-Bot or similar assistive robot should be treated as a serious occurrence, both in terms of the risks presented by inherent characteristics of the robot and the risks presented by the environmental situation at the time of failure.

B. Functional hazards

Software failure is a primary cause of functional hazards in the Care-O-Bot, as it can result in behaviours being carried out incorrectly or not at all. Software failure has been extensively studied in complex systems [12], and methods for assessing the contribution of development techniques to safety [13] are common across multiple domains. In addition, existing research has examined the correlation between failure rate estimates and verification performed [14].

However, a significant complexity for assistive robots such as the Care-O-Bot is the ability for end-users to define their own desired robot behaviours. Because of this, it cannot be assumed that the safety-critical behaviours of an assistive robot are known at the time of deployment. Notwithstanding verification such as [7], there is the potential for an inexperienced end-user to define behaviours which impact safety, or which put the robot in a position which can violate assumptions about the constraints it will obey. For example, an inexperienced user may define a behaviour which causes the robot to remind them to take their medication at an incorrect period or frequency. Equally, a user may define a behaviour which causes the robot to remain in another room, thus compromising its availability to perform those safety-critical functions which rely on direct observation of the user.

As with all systems, there is a UK legal requirement that the risk posed by assistive robots should be reduced As Low As Reasonably Practicable (ALARP) [9]. This requires hazards to be identified, risks to be estimated and mitigation to be put into place where needed to reduce the system risk to a tolerable level. In the case of assistive robots, the robot itself is typically taking a monitoring role and acting as partial mitigation for a wider risk. For example, a robot programmed to remind the user to take medication is partially mitigating against the illness which will result from a lack of medication. Similarly, a robot programmed to notify the user if the oven has been left on is partially mitigating against the risk of fire.

In each of these cases the user is required to take action to complete the mitigation (take the medication, switch off the oven, or evacuate the home). This is an effect of the fundamental design principles of the robot, driven by the need to prioritise *reablement*[2]. Reablement is defined as the drive to "Support people to do rather than doing to / for people"

 ISBN: 978-1-61208-698-9

[15] and is an important characteristic for service and assistive robots. Designing with reablement in mind means that the assistive robot is not intended to carry out the tasks itself (e.g., administering medicine to a user), but is instead intended to encourage the user to complete the task themselves. A side-effect of this design principle is that an assistive robot will typically require human engagement in order to successfully mitigate safety risks by completing the necessary action. One of the most important aspects of safety-critical assistive robot performance is therefore determined by the extent to which end-users engage with the robot.

IV. Socially appropriate behaviours

Because assistive and service robots are used within a domestic environment, it is important that the behaviour they display is both empathic and socially interactive [5]. Specifically, the behaviour a robot exhibits must be appropriate to the social role that it is expected to fulfil [16]. The extent to which a robot exhibits socially appropriate or socially intelligent behaviour is characterised by a number of factors, including its ability to establish and maintain social relationships, use natural cues, and express and perceive emotions [17].

Much existing work has explored the viability of transfering models of human interaction to robots, including an examination of adequate interaction distances and orientations [16], [18], [19]. Pursuing a complementary approach, research into the Care-O-Bot [6] has also exploited many techniques from the "learning by following" model [20]. Under this model the robot learns desired behaviours from following, observing and interacting with the human. The robot also conveys its capabilities and intentions using social signals [21] that might involve using whole-body or arm movements.

The social norms relevant to the robot will vary with its environment and operational use. Some may be generalised to a certain extent, modulo cultural differences. It is likely that there are certain situations in which it would be inappropriate for the robot to follow the user or capture their attention. For example, the human may have expectations of privacy which would be violated by the robot following them into the bathroom or bedroom [22][23]. Similarly, the human may have the social expectation that when they're engaged in a particular task (e.g., conducting a conversation), that the robot will not interrupt. Other social norms relevant to a domestic environment include detecting and adapting to a user's personal space, involving the user in decisions about entertainment and companionship and respecting the user's autonomy. Social norms will vary depending on the level of care required by the user, the degree of autonomy they expect, their age and personal preferences for interaction, as well as existing wider cultural and social constraints.

A. Social credibility

In this paper we extend the notion of socially appropriate behaviour to encompass the concept of *social credibility*. The social credibility of a domestic robot is a measure of how well it obeys the social norms relevant to its environment.

Social credibility helps determine the extent to which a human considers the robot to be a functioning social being. Work in [24] demonstrates a link between social intelligence and consideration of the robot as a social being. Further experiments have reinforced this tendency of humans to treat a socially intelligent or emotionally empathic robot as a social being, even to the extent of exhibiting concern over "hurting its feelings" [25]. This is amplified in a domestic or home setting, with end-users asked to rank the utility of cleaning robots considering their emotional impact as well as their functionality [26].

Social credibility has both a static and dynamic element. The static element refers to design: Has this robot been designed to follow social norms? Are its behaviours consistent with its appearance so that both match a potential user's expectations [21]? Static social credibility is also achieved via constraints embedded within the robot's programming (e.g., "do not follow a human into the bathroom").

Dynamic social credibility refers to the ongoing adaptability of the robot's behaviour: is it capable of adjusting its own behaviours based on feedback and the observed environment? Dynamic social credibility allows for evolution of the social norms over time. For example, it may be within norms for a domestic robot to follow a child user into a bedroom, but not for it to similarly follow an adult user. As a child user ages, dynamic social credibility ensures that the robot's behaviour reflects the changing application of the norm.

Social credibility is an evolving measure, and dependent on the actions of the robot. Much as a system which does nothing is "perfectly safe", a robot which is turned off and hence never takes an action will not lose nor gain social credibility. Social credibility may be temporarily lost by an inappropriate action, and gained back by subsequent actions.

As discussed in Section IV, social norms will vary with the environment. For a domestic service or assistive robot, we consider the following contributors (both positive and negative) to social credibility:

- Frequency and urgency of interruptions
- Nature and intensity of interaction, engagement and interruption
- Responsiveness of the robot to verbal and non-verbal feedback
- Appropriate physical movement and distance maintained from end-user
- Trust inspired by the robot in the end-user
- Understanding communicated by the robot as to its capabilities

It is important to note that although trust is a significant aspect of social credibility for an assistive robot, it is not the only factor. Much work already exists on the questions of eliciting and maintaining trust (see [27] for an overview, additionally [28][29]), with considerations of factors such as reliability, predictability, physical presence and emotional response.

However, it is possible for a robot to inspire trust and emotionally engage a user without necessarily having a high degree of social credibility. For example, a pet-like robot [30] may emotionally engage a user because of its appearance and actions, but there are typically fewer social norms applicable to a pet. Similarly, an autonomous vehicle or an alarm system may be trusted by its end users without any imputation of sociability or social knowledge. By contrast, a robot which shares personal information about its user with a third party will typically be regarded as untrustworthy [31], but such sharing does not in

 ISBN: 978-1-61208-698-9

itself mean the robot is not seen as a social being (a malicious person may have also done the same).

Crucially, social credibility also requires that the user understand the robot's capabilities, much as they would understand the different capabilities of a human adult or a human child. A high degree of social credibility implies that a robot has communicated an understanding as to its capabilities and reduces the potential for over-trust [32]. From the perspective of safety, over-trust is considered a negative factor as it leads to excessive reliance on the automation even when there are indications of system failure.

V. Social credibility and safety-critical systems

A large part of the duties of an assistive robot involve reminding or prompting the end-user to take action. This involves some form of interruption to the user's current activity. Two important social norms for these robots are therefore around the frequency of interruptions and on the way these interruptions are made. In [33], users explicitly identify that reminders given by an assistive robot become irritating under the following circumstances:

- When repeated often
- When repeated in a "mechanical" voice
- When repeated at inopportune times, interrupting the user

Conversely, some behaviours and methods of interruption are viewed positively by users and considered to mimic human interruptions, as discussed in [19] and [34]. These include the use of direct, random and non-random gaze directions to signal the beginning of an interaction. Other studies [35][36] have examined users' preference for personal space from robots, identifying that users perception of their personal space diminishes for likeable robots, and similarly that robots which encroach on this space are regarded as unlikeable, threatening or irritating. Personal space preferences will vary with context; for example, users are typically reluctant to accept a robot following them into the bathroom [22].

Inappropriate interruptions therefore present a potential for a loss of social credibility. A robot whose interruptions take no account of social norms is more likely to be regarded as a simple mechanical system (e.g., an alarm or reminder application) instead of as another social entity. For example, an assistive robot which always sounds an alarm at a certain time to remind the user to take medication is performing a role no more complex than an alarm clock, and hence complying with no relevant social norms. As such, it does not build social credibility in the same way that an assistive robot would if its interruptions were sensitive to the users' environment, engagement and current activities ([19]). As social credibility is a dynamic concept (see Section IV-A), a robot which has already built social credibility by demonstrating such sensitivities is vulnerable to losing this credibility if its interruptions become inappropriate.

A loss of social credibility (from any cause) can lead to an end-user disengaging with the robot in a number of ways. Firstly, the user may simply switch the robot off. Studies have shown that users are reluctant to switch off robots they consider to be intelligent [37], or perceived social beings. However, once social credibility is lost, this "protective" aspect is lost with it. Users are much more willing to switch off a robot considered to be solely a *robotic device*, particularly when the mode of engagement with this robot becomes arduous. In [38], drivers concluded that they would prefer to be able to turn off a speed warning system that was judged "irritating", even where they agreed that use of the technology would be helpful.

Secondly, even where the user permits the robot to remain switched on, they may start to ignore the suggestions and prompts made by the robot. This then leads to a dilemma for those designing such robots: if repeated interruptions lower social credibility, then how should the robot deal with an urgent prompt that has been ignored?

A. Safety-critical systems

Any disengagement with an assistive robot (whether switching it off or ignoring its prompts) compromises its ability to perform its safety-critical functions. It is clear that switching a robot off renders it incapable of providing any alerts or reminders. Similarly, because assistive robots mitigate risk by prompting end-user action (see Section III), any user disengagement means that the risk mitigation is not carried out in full. For example, a robot reminding the user that the oven has been left on has no effect unless the user engages with the robot, and returns to switch the oven off.

Furthermore extrapolating from studies performed in other domains has enabled us to identify a unique user reaction that may result from loss of social credibility, and which affects only safety-critical actions of the robot (as opposed to routine actions). In more detail, safety-critical situations are the exception, not the rule, and hence any alert or reminder in such a situation will be perceived by the user as "not the expected behaviour". In the aviation domain, where autonomous cockpit systems are not considered to be social entities, pilots have been observed to attempt to debug the automation when its actions deviate from those they expected. In a study of cockpit automation [39] established the tendency in pilots to monitor automation status via the flight control unit (FCU), which shows *commanded* targets, paths and modes, rather than via the display showing *actual* targets, paths and modes being executed by the automation.

Given that this observation took place in a highly-trained cohort of pilots, it is reasonable to say that untrained end-users of an assistive robot may also display the same mode confusion. This would lead to a situation in which a user is alerted to a hazard and instead of taking mitigating action attempts to debug or force the assistive robot to return to the "expected" behaviour.

B. Prioritising safety-criticality

The performance of safety critical behaviours is a clear priority from a legal and regulatory viewpoint [9][8] (for other priorities, see V-C). Motivated by this, we have identified a number of potential methods to address loss of safety-critical functionality resulting from lowered social credibility. Each of these methods trades a slight decrease in the robot's overall capability in return for maintaining an adequate level of social credibility. Since social credibility is a requirement for effective safety critical performance, this corresponds to decreasing the robot's capabilities in order to gain confidence that safety-critical engagements will be performed effectively when needed.

 ISBN: 978-1-61208-698-9

The first method we describe is an attempt by the robot to alter its behaviour when the social credibility drops below a threshold value which we will term the *disengagement threshold*. The disengagement threshold is the level of social credibility at which engagement with the robot (including its future safety-critical behaviours) is jeopardised. When this threshold is being approached, the robot should choose to alter the nature of its alerts and reminders to stop social credibility loss.

Both [19] and [34] identify a number of methods whereby a robot may interrupt an end-user, based on non-verbal behavioural cues. The extent and urgency of the interruption can be tailored to its nature: a safety-critical behaviour may still merit an urgent (and socially inappropriate) interruption even when the robot's social credibility is at risk of dropping below the threshold. However, for less critical interruptions the robot may choose to utilise any of the following behaviours:

- Slow its physical movements when coming to interrupt a user
- Decrease the volume of any audible alerts
- Display visual alerts (e.g., on the attached screen, for the Care-O-Bot), instead of audible alerts
- Approach the user and wait for the user to initiate an interaction

In addition to altering the nature of its alerts and interruptions, the robot may also choose to alter the frequency of these when approaching the disengagement threshold. Interruptions are a cognitive challenge for a user, and existing work shows that in some situations user satisfaction is maximised by delaying an interruption at the cost of some awareness [40].

This proposal allows a robot to *delay* a routine behaviour (such as interrupting the user with the offer of food or drink) in order to retain sufficient social credibility to ensure that any safety-critical behaviour (such as notification the oven is on) will be engaged with by the user. Other routine behaviours a robot may choose to suspend or delay if its social credibility is low include: greeting the user, engaging in social interaction and conversation, reminding the user of appointments and offering the user entertainment.

C. Prioritising social credibility

However, safety-critical performance is not the only consideration for assistive robots. It is also imperative that these robots perform their social functionality adequately. There is the potential for prioritisation of functionality relating to safety (e.g., requiring the robot to follow the user through the house in case of a fall) to result in the neglect of other socially important behaviours such as greeting, user engagement and user interaction. In other words, a robot performing only safety-related behaviours may not be free to perform other roles which are critical to its reablement functionality.

Furthermore, the performance of the safety-critical behaviours can itself lead to a loss of social credibility. A robot alerting the user to a fire may out of necessity do so at an inopportune time or in an urgent or disruptive fashion. The nature of such (intense, potentially ill-timed) alerts means that they will result in a certain loss of social credibility. This has the potential to drive the social credibility of the robot below the disengagement threshold, and therefore result in reduced capability (both routine and safety-critical) due to lack of user engagement.

A loss of social credibility has significant impact on the socially important aspects of a robot's functionality. In more detail, the characteristics identified in Section IV-A as associated with social credibility are also important for user engagement. Trust, for example, means that a user is likely to extrapolate from observed characteristics of the robot to generalise about its wider capabilities [41]. While over-trust is in itself a problem [28], a lack of trust in the robot means that users are likely only to engage the robot in scenarios which they have directly observed to be satisfactorily carried out. That is, even where the overall social credibility has not been driven below the disengagement threshold, the overall capability of the robot may still be impaired. Similarly, a lack of trust in a robot may lead to negative associations with it, and a reluctance on the part of the user to engage [29].

It is therefore clear that a balance will need to be struck between performing safety-critical behaviours, and performing the social routines necessary to build user engagement (socially-important behaviours).

D. Schedulability of behaviours

We propose the identification of an optimum scheduling such that socially-important and safety-critical behaviours can both be performed to an acceptable level. This will correspond to maintaining social credibility above the disengagement threshold by delaying behaviours based on their priority, where priority considers both safety and social engagement.

Such a prioritisation system would correspond to trading off (safety) risks against (social) benefit, a concept described in [42]. Traditional scheduling algorithms could be used to ensure that the correct behaviours are selected to run, with a level of customisation also being provided.

This problem has been explored extensively when considering scheduling within mixed-criticality systems (see [43] for an overview). In the case of assistive robots, the following (non-exhaustive) criteria should be considered for schedulability:

- Estimated risk associated with not fulfilling the behaviour
- Estimated loss of social credibility associated with fulfilling the behaviour
- Current social credibility as considered against the disengagement threshold
- Functional importance of other behaviours

Such a prioritisation system could also be customised to allow users and care-givers to adjust the balance between safety and social behaviours. A user more comfortable and engaged with the robot may not need the same degree of social behaviours as a user who has not engaged with the robot before. Similarly, a user requiring a higher level of care may want to prioritize safety-critical behaviours.

VI. Conclusions and Future Work

In this paper we have identified a link between safety and *social credibility*, where this is defined as being a reflection of how well a robot follows social norms. The relevant social norms are dependent on the environment and purpose of the robot, and we have presented some examples that would apply

 ISBN: 978-1-61208-698-9

to an assistive robot. We have also drawn on analysis of existing systems to identify how user disengagement can affect both social credibility and the safety-critical functions of an assistive robot. In this process, we have shown how loss of social credibility can lead to effective loss of these safety functions.

We have built on this in order to discuss prioritisation of socially-important behaviours and safety-related behaviours, particularly where these may conflict. Over-prioritisation of safety-related behaviours can itself lead to a loss of social credibility, and to user disengagement. Correspondingly, over-prioritisation of routine behaviours can lead to poor performance in the robot's safety-related roles. We have proposed a solution to this that builds on existing concepts of mixed-criticality system scheduling. Such a scheduling would rely on a prioritisation system that takes both safety and social engagement into account.

As part of future work, we propose to develop this prioritisation further. We will evaluate in a user study how exactly social credibility could be affected by violations of social norms that are required from a safety point of view. Furthermore, we plan to investigate how safety-relevant routines might be neglected by the user when the robot is not perceived as socially credible. This data will be used in studies futher investigating the automatic scheduling of behaviours to ensure the robot maintains high levels of social credibility while being acceptably safe to operate. We also propose to expand this work to discussions of other autonomous systems, providing a generalised mechanism for assuring safety of a robot which must also perform another (social) function.

REFERENCES

[1] J. Broekens, M. Heerink, and H. Rosendal, "Assistive social robots in elderly care: A review," *Gerontechnology*, vol. 8, no. 2, pp. 94–103, 2009.

[2] F. Amirabdollahian *et al.*, "Assistive technology design and development for acceptable robotics companions for ageing years.," *Paladyn: Journal of Behavioral Robotics*, vol. 4, no. 2, pp. 94–112, 2013.

[3] R. Kittmann *et al.*, "Let me introduce myself: I am care-o-bot 4, a gentleman robot," in *Mensch und Computer 2015 – Proceedings*, S. Diefenbach, N. Henze, and M. Pielot, Eds., Berlin: De Gruyter Oldenbourg, 2015, pp. 223–232.

[4] J. Saunders, N. Burke, K. L. Koay, and K. Dautenhahn, "A User Friendly Robot Architecture for Re-ablement and Co-learning in A Sensorised Home," in *European AAATE (Associated for the Advancement of Assistive Technology in Europe) Conference*, Vilamoura, Portugal, 2013, pp. 49–58.

[5] S. Bedaf, P. Marti, F. Amirabdollahian, and L. de Witte, "A multi-perspective evaluation of a service robot for seniors: The voice of different stakeholders," *Disability and Rehabilitation: Assistive Technology*, pp. 1–8, 2017.

[6] J. Saunders, D. Syrdal, K. L. Koay, N. Burke, and K. Dautenhahn, "'Teach Me - Show Me' - End-user personalisation of a smart home and companion robot," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 1, pp. 27–40, 2016.

[7] M. Webster *et al.*, "Toward reliable autonomous robotic assistants through formal verification: A case study," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 2, pp. 186–196, 2016.

[8] International Standards Organization, "Robots and robotic devices - safety requirements for personal care robots," *ISO 13482*, 2014.

[9] UK Health and Safety Executive, *Reducing Risks, Protecting People*. HSE Books, London, UK, 2001.

[10] Fraunhofer IPA, *Care-o-bot data sheet*, 2018.

[11] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *IEEE Robotics & Automation Magazine*, vol. 4, no. 1, pp. 23–33, 1997.

[12] J. C. Knight, "Safety critical systems: Challenges and directions," in *Proceedings of the 24th International Conference on Software Engineering*, ACM, 2002, pp. 547–550.

[13] International Electrotechnical Commission, "Functional safety of electrical / electronic / programmable electronic safety related systems," *IEC 61508*, 2010.

[14] J. McDermid and T. Kelly, "Software in safety critical systems-achievement & prediction," *Nuclear Future*, vol. 2, no. 3, p. 140, 2006.

[15] UK Department of Health, *Care services efficiency delivery programme 'homecare re-ablement workstream: Retrospective longitudinal study'*, 2007.

[16] K. L. Koay, D. S. Syrdal, M. Ashagari-Oskoei, M. L. Walters, and K. Dautenhahn, "Social Roles and Baseline Proxemic Preferences for a Domestic Service Robot," *International Journal of Social Robotics*, vol. 6, no. 4, pp. 469–488, 2014.

[17] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robotics and autonomous systems*, vol. 42, no. 3-4, pp. 143–166, 2003.

[18] D. S. Syrdal, K. Dautenhahn, M. L. Walters, and K. L. Koay, "Sharing Spaces with Robots in a Home Scenario – Anthropomorphic Attributions and their Effect on Proxemic Expectations and Evaluations in a Live HRI Trial," in *AAAI Fall Symposium "AI in Eldercare: New Solutions to Old Problems"*, Washington, DC, USA, 2008, pp. 116–123.

[19] P. Holthaus, K. Pitsch, and S. Wachsmuth, "How Can I Help? - Spatial Attention Strategies for a Receptionist Robot," *International Journal of Social Robotics*, vol. 3, no. 4, pp. 383–393, 2011.

[20] M. N. Nicolescu and M. J. Mataric, "Learning and interacting in human-robot domains," *IEEE Transactions on Systems, man, and Cybernetics-part A: Systems and Humans*, vol. 31, no. 5, pp. 419–430, 2001.

[21] F. Hegel, S. Gieselmann, A. Peters, P. Holthaus, and B. Wrede, "Towards a typology of meaningful signals and cues in social robotics," in *2011 RO-MAN*, 2011, pp. 72–78.

[22] D. Feil-Seifer and M. J. Matarić, "Socially assistive robotics," *IEEE Robotics & Automation Magazine*, vol. 18, no. 1, pp. 24–31, 2011.

[23] D. T. Anderson, J. M. Keller, M. Skubic, X. Chen, and Z. He, "Recognizing falls from silhouettes," *Electrical and Computer Engineering publications (MU)*, pp. 6388–6391, 2006.

[24] M. Heerink, B. Kröse, V. Evers, and B. Wielinga, "The influence of social presence on acceptance of a companion robot by older people," *Journal of Physical Agents*, vol. 2, no. 2, pp. 33–40, 2008.

[25] A. Hamacher, N. Bianchi-Berthouze, A. G. Pipe, and K. Eder, "Believing in bert: Using expressive communi-

 ISBN: 978-1-61208-698-9

cation to enhance trust and counteract operational error in physical human-robot interaction," in *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*, IEEE, 2016, pp. 493–500.

[26] J. Forlizzi, "How robotic products become social products: An ethnographic study of cleaning in the home," in *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, ACM, 2007, pp. 129–136.

[27] J. Lee and K. See, "Trust in automation: Designing for appropriate reliance," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 46, pp. 50–80, 2004.

[28] M. Salem, G. Lakatos, F. Amirabdollahian, and K. Dautenhahn, "Would You Trust a (Faulty) Robot?: Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust," in *International Conference on Human-Robot Interaction (HRI)*, Portland, Oregon, USA: ACM/IEEE, 2015, pp. 141–148.

[29] M. Desai, K. Stubbs, A. Steinfeld, and H. Yanco, "Creating trustworthy robots: Lessons and inspirations from automated systems," in *Proceedings of the AISB Convention: New Frontiers in Human-Robot Interaction*, 2009, pp. 49–56.

[30] C. D. Kidd, W. Taggart, and S. Turkle, "A sociable robot to encourage social interaction among the elderly," in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, IEEE, 2006, pp. 3972–3976.

[31] A. Rossi, K. Dautenhahn, K. L. Koay, and M. L. Walters, "How the timing and magnitude of robot errors influence peoples' trust of robots in an emergency scenario," in *International Conference on Social Robotics*, Springer, 2017, pp. 42–52.

[32] M. Salem, G. Lakatos, F. Amirabdollahian, and K. Dautenhahn, "Towards Safe and Trustworthy Social Robots: Ethical Challenges and Practical Issues," in *International Conference on Social Robotics*, Cham: Springer International Publishing, 2015, pp. 584–593.

[33] S. Bedaf, H. Draper, G.-J. Gelderblom, T. Sorell, and L. de Witte, "Can a service robot which supports independent living of older people disobey a command? the views of older people, informal carers and professional caregivers on the acceptability of robots," *International Journal of Social Robotics*, vol. 8, no. 3, pp. 409–420, 2016.

[34] P. Saulnier, E. Sharlin, and S. Greenberg, "Exploring minimal nonverbal interruption in hri," in *RO-MAN, 2011 IEEE*, IEEE, 2011, pp. 79–86.

[35] A. Sardar, M. Joosse, A. Weiss, and V. Evers, "Don't stand so close to me: Users' attitudinal and behavioral responses to personal space invasion by robots," in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, ACM, 2012, pp. 229–230.

[36] J. Mumm and B. Mutlu, "Human-robot proxemics: Physical and psychological distancing in human-robot interaction," in *Proceedings of the 6th international conference on Human-robot interaction*, ACM, 2011, pp. 331–338.

[37] C. Bartneck, T. Kanda, O. Mubin, and A. Al Mahmud, "Does the design of a robot influence its animacy and perceived intelligence?" *International Journal of Social Robotics*, vol. 1, no. 2, pp. 195–204, 2009.

[38] J. Wall, V. Cuenca, K. Creef, and B. Barnes, "Attitudes and opinions towards intelligent speed adaptation," in *Intelligent Vehicles Symposium Workshops (IV Workshops), 2013 IEEE*, IEEE, 2013, pp. 37–42.

[39] N. B. Sarter and D. D. Woods, "Team play with a powerful and independent agent: Operational experiences and automation surprises on the airbus a-20," *Human factors*, vol. 39, no. 4, pp. 553–569, 1997.

[40] E. Horvitz, J. Apacible, and M. Subramani, "Balancing awareness and interruption: Investigation of notification deferral policies," in *International Conference on User Modeling*, Springer, 2005, pp. 433–437.

[41] P. Robinette, W. Li, R. Allen, A. M. Howard, and A. R. Wagner, "Overtrust of robots in emergency evacuation scenarios," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, IEEE Press, 2016, pp. 101–108.

[42] C. Menon and R. Alexander, "Ethics and the safety of autonomous systems," in *Proceedings of the 26th Safety Critical Systems Symposium*, Safety Critical Systems Club, 2018, pp. 25–43.

[43] A. Burns and R. Davis, "Mixed criticality systems-a review," *Department of Computer Science, University of York, Tech. Rep*, pp. 1–69, 2013.

 ISBN: 978-1-61208-698-9