Social Assistive Robots, Reframing the Human Robotics Interaction Benchmark of Social Success

Antonio Espingardeiro

Abstract—It is likely that robots will cross the boundaries of industry into households over the next decades. With demographic challenges worldwide, the future ageing populations will require the introduction of assistive technologies capable of providing, care, human dignity and quality of life through the aging process. Robotics technology has a high potential for being used in the areas of social and healthcare by promoting a wide range of activities such as entertainment, companionship, supervision or cognitive and physical assistance. However such close Human Robotics Interaction (HRI) encompass a rich set of ethical scenarios that need to be addressed before Socially Assistive Robots (SARs) reach the global markets. Such interactions with robots may seem a worthy goal for many technical/financial reasons but inevitably require close attention to the ethical dimensions of such interactions. This article investigates the current HRI benchmark of social success. It revises it according to the ethical principles of beneficence, non-maleficence and justice aligned with social care ethos. An extension of such benchmark is proposed based on an empirical study of HRIs conducted with elderly groups.

Keywords—HRI, SARs, Social Success, Benchmark, Elderly care.

I. INTRODUCTION

According to United Nations (UN) there is an emerging ageing phenomenon worldwide [1]. It is likely that ageing populations will require extra levels of physical and cognitive assistance throughout their lives. A great deal of attention must be directed to assistive technologies for facilitating ageing-in-place, living independently and promoting the wellbeing of individuals and communities. Robotics as a multidisciplinary science starts to demonstrate potential to be used in social care contexts [2]. The first generation of Social Assistive Robots (SARs) are likely to deliver cognitive assistance, supervision, entertainment and even companionship for elderly groups. SARs developments should be targeted to complement elderly care and be used as an extension of caregiving activities. At the same time the introduction of SARs is likely to raise ethical challenges around independency versus human contact, privacy and wellbeing of elderly groups. Currently there is a paucity of studies that involve the use of SARs with vulnerable groups such as the elderly. As an example existing studies highlight psychological gains with the use of SARs with elderly groups but also report emerging connections formed between elderly groups and robotic animals [3], [4]. Similarly [3] highlights positive results in terms of communication and socialization of elderly groups with the use of robotic seals but also report emerging connections between certain individuals and such robots. However none of these studies analyses the use of SARs from an ethical standpoint. It is understandable that further ethical analysis is required to help understanding some of the emerging ethical issues arising from the use of SARs in elderly care. Additionally we will need the development of Roboethics frameworks that can help developers and potential users to interact with the first generation of SARs.

II. ROBOETHICS

According to [5], Roboethics can be defined as a new field of study that “tries to develop scientific, cultural, technical tools for deploying robots into a wide range of social groups and believes”. Because SARs philosophy is focused on the outcome of HRI in this article we will explore the existing HRI benchmark of “Social Success” [2]. We will try to refine it based on practical robotics workshops with observations and interviews/comments from elderly residents in care/extra care homes. The study of HRI benchmarks constitute an important contribution for guiding robotics developers and other stakeholders to understand some of the emerging ethical issues arising from the introduction of the first generation of SARs.

To date robots with various forms and dimensions are being equipped with sensors and computerized with artificial intelligence algorithms for a wide range of purposes [6]. According to some of the world leading experts in robotics [7], [8] it is likely that robots will be endowed with the ability to learn and process human profiles, tastes, habits, which will inevitably lead to privacy, safety and individual freedom choices. If that is the case humans will coexist with a new generation of automated machines (robots) employed as domestic workers, nurses and caregivers at home, hospitals and extra care facilities. When such scenario will take place it is likely that this widespread distribution of robots will raise several completely new ethical, legal, and societal challenges. Roboethics constitutes an area of primordial importance to be studied and further developed. According to [9] robots can originate psychological issues, especially in vulnerable groups such as children, elderly people and hospital patients. Also issues regarding the attribution of civil and criminal liability should an autonomous robot produce damages are arising in many debates. Such ideas have been subject to discussion since the “dawn” of robotics in the works of [10] or in the science fiction stories of Isaac Asimov such as “Runaround” [11] where “the three laws of Robotics” were introduced. However according to [12] Asimov laws were conceived.
purely for science fiction scenarios. The laws were derived from a top-down approach where the ethical theories of utilitarianism and deontology were applied. The laws seem to imply that robots have similar cognitive abilities and behaviours as human beings. However such laws are far from being implemented into robots due to the current state of the art of robotics [12] and the lack of understanding of the human brain and consciousness (how does it work and how can it be implemented on robots?).

It was only in the last few years that the debate has been progressively organized within the international robotics community and that the term “Roboethics” has established itself as an emerging field of applied ethics.

Veruggio [6] classifies Roboethics into three distinct levels: Roboethics level 1 which can be understood as an ethical reflection directly related to the particular issues that are generated by the development of robotic applications and their diffusion in society. Robot Ethics level 2 which regard the code of conduct that developers should consider in artificial intelligence algorithms of their robotic creations. Finally Veruggio presents Robot’ Ethics level 3 as a hypothetical scenario where robots were equipped with conscience and freedom to choose its own actions on the basis of a full comprehension of their acts and consequences. Beyond the existing classification and definition it’s likely that HRIs will open a new set of ethical challenges, experiences, behaviours and relationships between humans and machines.

III. ROBOT EVALUATION, HRI BENCHMARKS

It seems plausible that SARs as technological platforms for human assistance must be evaluated before they are deployed within the proximity of humans or the environment. Along those lines existing works around SARs developments are reviewed. Previous works [13]-[16], [2] discuss the topic of SARs developments and potential use with vulnerable groups but do not propose frameworks/tools or guiding steps towards the development of SARs. As the foundational definition of [6] states Roboethics tries to “develop scientific, cultural, technical tools for deploying robots into a wide range of social groups and believes”. Such exercise represents an attempt to guide and conceive ways of producing better SARs for complementing human beings. Feil-Seifer et al. [2] draws more attention in terms of SARs developments and potential guidance. In the area of robotics technology (Table I) [2] identifies two benchmarks “Safety” and “Scalability” whereas in the social interaction domain “Autonomy”, “Imitation”, “Privacy”, “Understanding of Domain” and “Social Success”.

However, such benchmarks are mainly inspired by psychology and do not contemplate an ethical analysis on its core development. In the domain of elderly care it is believed that SARs could become extremely important in the exercise of care, however we need further discussion to better inform future developments and introduction of SARs. Equally important is the need for more field work related with care and extra care facilities where real HRIs could take place. Social care ethos has a determinant role in listening to people’s perspectives, expectations, dignity and choices throughout their care. In practical terms we find evidence that the core ethical principles and their application in caring for older people present big challenges. Suhonen et al. [18] identify ethically difficult situations in the care of older people where there is evidence that perceptions differ about ethical issues among health professionals, patients and their relatives. The application of the core ethical principles could have different interpretations within different contexts of application. Scott et al. [19] conclude that improvements in nursing care for elderly people seem to demand greater levels of communication between caregivers and care receivers. Communication could help ensure that the staff teams have a better understanding of what information and what level of involvement in decision making regarding care, patients need or want.

At this point it is understandable that we must enrich the current HRI benchmarks knowledge with both an interpretation of the core ethical principles but also considering social care ethos. Over the next sections we will be analysing the HRI benchmark of social success. We will be reviewing such benchmark in line with the commonly accepted core medical ethics principles of beneficence, non-maleficence and justice [20] aligned with social care ethos. At this stage we haven’t considered the ethical principle of autonomy as SARs are in testing phases. However it is important to recognize that whatever results emerge from the social success benchmark could significantly influence the elderly decision about the inclusion of robots in their own care.

IV. SOCIAL SUCCESS

According to [2] the task oriented benchmark of social success (Table I) tries to understand if SARs accomplish their primary objectives. As an example if a robot is programmed to being funny, is it really being funny? However, in ethical terms such vision might be too reductionist. Initially SARs should be designed for promoting the wellbeing (beneficence) of elderly individuals, but for example the relation between SARs success and the ethical principle of non-maleficence is extremely complex. If we consider examples where robotic animals are used as relaxation exercises to comfort elderly people in care homes the notion of success could become relative. Academic studies [21], [22], [4] refer to notions of attachment taking place between vulnerable groups and robotic animals such as PARO (a baby robotic seal used in care homes in Japan and USA). When attachment takes place the phenomenon could be seen as an excess of success. However the opposite could also happen. Deception could
occur in other social robotics scenarios when for example a robot doesn’t meet its user expectations in HRIs. To date the psychological repercussions of such phenomena in elderly groups is still unknown. However information about the robots capabilities and direct behavioural responses are extremely important to be clarified. As SARs have a synthetic appearance and since humans are heavily influenced by visual cues, we could expect several types of instant responses to robot appearance [23]. The notion of scale (size of robot), the concept of usability (how to turn it on/off, how to interact with it), or even the way the robot is “dressed” and accessorized could influence the way it is perceived by elderly groups. It is highly probable that social care ethos will play an important role in determining or not the success of HRIs. As a result personalizing elements in HRIs could arise and will need to be identified as they can positively inform future SARs developers and manufacturers.

Still in non-maleficence there is the notion of meaning and earnestness. High levels of HRI could also translate false expectations when for example a vulnerable user communicates health problems to robot and expects it to inform an agency (health care) or react like a real clinician. Sensitive information about a person’s health and wellbeing might fall into such scenario that can originate ethical repercussions. To aggravate such challenge is the fact that the loss of earnestness and machine authority during HRIs may not be instantaneous. The user might be receptive and amenable to interact with a SAR for some initial period perhaps due to the novelty of the machine, however the user might lose interest in it with time [24]. So to act in ways that both benefit and do not harm users, SARs systems should be constantly updated to create high expectations throughout the interaction life cycle. However the solution for such issues isn’t likely to emerge solely from algorithms and robotic behaviours. We might need further engagement of caregivers, relatives, users and robots to continuously cultivate meaning to HRIs through classical social interaction. Lastly the ethical principle of justice talks about the fair distribution of resources. If SARs are going to be implemented in the near future then care institutions have to debate the fair access to such type of technology, how to supervise HRIs, maintenance of SAR systems and responsibility towards them. Beyond the access challenge, justice also questions the benefit and cost of such HRIs which could become inspired by existing governmental health systems policies across nations.

Such types of researches and clarifications are expected to be challenging with vulnerable groups that frequently suffer from cognitive problems. Questions such as where is the boundary between comforting exercises and addiction to robots in elderly groups? How to act in cases of robotic attachment or losses of interest? What is the responsibility of caregivers and clinicians relative to such types of practices, and where is the line between living more independently and becoming socially isolated? All SARs four core areas of supervision, cognitive assistance, entertainment and companionship pose similar challenges that need to be further analysed. Social care ethos will involve talking to elderly groups to analyse their perspectives, attitudes, dignity and expectations towards social success in SARs.

In terms of the ethical principles selection in the benchmark of social success we are primarily concerned with the qualitative elements that can build good levels of HRIs. Thereby we are considering the ethical principle of beneficence as the HRIs should be constructed for the benefit of elderly groups. On the same line the ethical principle of non-maleficence is important to avoid potential situations where HRIs could possibly harm elderly individuals. As social success is researched a fundamental question arises with the fair access and distribution of SAR technologies that can benefit elderly groups. Thereby the ethical principle of justice should be considered. Lastly as social success represents a set of qualitative elements also elderly groups’ opinions and expectations towards SARs are crucial to be analyzed. So, social care ethos is crucial here. In the benchmark of social success we are considering the ethical principles of beneficence, non-maleficence and justice aligned with social care ethos.

V. REFRAMING APPROACH

Robotic workshops were prepared for interaction with elderly groups. The experiments took place in 5 institutions in the UK and Portugal. The study had 74 elderly participants plus caregivers, institutional managers and relatives. We conducted in-situ research which meant a richer set of observations to be registered and post analyzed. The research methodology employed was an interpretivist philosophy using qualitative methods that involved observations, interviews and informal comments analysis.

The researcher and robots were performing in a common care/extra care environment (lounge) with the supervision of caregivers. The robots were controlled in real time by a researcher. The experiments were delivered as entertaining exercises taking place once per week in care institutions during a period of 7 months.

The robotic workshops involved 50cm humanoid robots programmed with songs and choreographies. Two mobile robotics platforms were used to demonstrate, supervision medication and tasks reminders routines. In a second activity robotic seals and robotic cats were also used as comforting and relaxation exercises with elderly groups.

In the following section a revised interpretation and categorization of Feil-Seifer's HRI benchmark of social success is proposed. Thereby the new interpretation results from a combination of the ethical analysis involving the ethical principles of beneficence, non-maleficence, justice aligned with social care ethos and a qualitative analysis resulting from practical robotic workshops with elderly groups [25]. Such process involved understanding the emerging results from the qualitative analysis but also revisiting the fundamental HRI benchmark of social success to refine and extend current knowledge on some of the ethical issues involved in the use of SARs with elderly groups. As a result we will revisit Feil-Seifer’s original HRI benchmark of social success and propose new categories.
VI. REFRAMED SOCIAL SUCCESS

In the social success benchmark we tried to understand potential qualitative elements that can build and reinforce the success of HRIs with elderly groups. The first point is to try to clarify what is the objective of such HRIs with elderly groups and what are the possible emerging questions (advantages and disadvantages) arising from those.

In terms of users' responses we started by analyzing if the elderly did prefer listening music from a robot or a classical radio [25]. Elderly groups did prefer listening music from a robot however issues were raised relative to the quality of the audio on the humanoid robot itself. An enquiry was also made relative to the use of more or less robotized voices. The elderly preferred the more robotized voice used in the humanoid robots.

Equally important was to understand the users' body language when the researcher gave and retrieve a ball from the robot in close proximity of the elderly. We found that the elderly were not afraid of the robots and were in fact supportive of close HRIs. In terms of personalization elements we did investigate if the elderly were supportive of uploading their favourite songs to the robots (or have someone that could do it for them). The response was positive. On the same line it is important to mention that ethnographic considerations did play an important role in defining the content to be programmed into the humanoid robots. Across the 5 different institutions investigations were made relative to language, songs and jokes that could be programmed into the robots. Such qualitative elements are likely to reinforce the outcome of the HRI. In terms of cognitive assistance we demonstrated potential scenarios where a SAR reminds the elderly about their medications and daily tasks. The elderly were supportive of such actions.

In social success we found that the notion of robotic presence could become determinant for the outcome of the HRI. In the D45 workshop (mobile robot) elderly participants were doubtful about the potential of such robot. D45 had no significant aesthetics work and didn’t had any anthropomorphic elements. In interview 3 comments were addressed “what strange machine is that”. It was clear that D45 didn’t achieve the notion of robotic presence among the elderly. Across the 5 different workshops investigations were made relative to language, confidence, level of communication and ethnographic studies and methods used to deliver SARs.

Questions:

Type of Robotic Application Delivered and Emerging Questions:

Initially it is important to clarify the type of robotic application used and what is the main objective in terms of HRI. This exercise is likely to reveal potential questions and answers that we want to expand through the form of existing HRI benchmarks. It seems the simple answer “yes” or “no” doesn’t include enough extension for understanding some of the emerging challenges of SARs.

Users’ Responses:

Social success in SARs has to try to explain why, how and when social success seems to be valid. Thereby the mechanisms by which we can qualitatively and quantitatively measure the results of HRI have to be yet researched. Such mechanisms could include analysing users’ responses in terms of body language, confidence, level of communication and socialization displayed during HRIs. It is important to stress
that independently from the level of autonomy displayed and autonomous supervision schemes there are several stakeholders involved in SARs (user, robot, human supervisor (caregiver)). In interview 2 [25] it is recommended that supervised HRI be analyzed in conjunction with an assessment panel which could be formed by e.g. researchers, staff and families. It is also important to retain the notion of content programmed and personalization in SARs. Such balance could make the HRI more or less successful. During interview 2 there were elements in HRIs such as colours or voices that could become personalizable and contribute for higher levels of immersion during HRIs.

Robotic Presence:

Robotic presence is a result of how well imitation is perceived within SARs however it is also dependent on the aforementioned human responses resulting from the robot’s behaviour. In elderly care, people are less likely to interact with SARs that do not transmit any sense of technological presence e.g. robots full of wires. This was particularly true in interview 3 [25] when D45 was demonstrated to the elderly groups. Comments were made "strange machine" or "are you sure it is safe?". Conveying robotic presence in SARs is equally related on how well the human machine interfaces are available to a user and the generic HRI experience is perceived.

Attachment:

Social success could become successful but also develop notions of attachment on individuals. During the robotics workshops we identified notions of attachment when it came to the robotic animals activities. Especially in interview 3 [25], elderly residents were constantly commenting “when we will have the robotic cats?” or “you can leave them with us”. In addition the elderly body language demonstrated high levels of connection with both seals and cats and in certain cases some of the female participants were reluctant to give the robots back.

Ethnographic Studies Informing SARs Content:

Social success also derives from the content programmed into a SAR. Thereby ethnographic studies could contribute to the overall result of SARs if there is affinity between man and machine.

Methods Used to Deliver SARs:

Lastly the methods used to conduct robotic activities with the participation of vulnerable groups have to be weighted also. Researcher and staff worked together (interviews 2 and 3) towards the social success of the robotic workshops [25]. The presenting methods seemed to work well with vulnerable groups. As an example theories of communication [26] and groups dynamics [27] became extremely relevant to read the audiences responses and to adapt the presenter scheme, skills and robot to deliver SARs with elderly groups.

VII. CONCLUSION

Social assistive robotics is a new area of research that is focused on the outcome of HRI in terms of assistance, rehabilitation, convalescence or learning. Robotics science starts to demonstrate a high potential for offering cognitive assistance, communication, supervision and entertainment for vulnerable groups. However the introduction of SARs within elderly communities is not an easy task. There are emerging ethical issues that must be explored through the use of HRI benchmarks for guiding robotics developers and ultimately users when it comes to develop and use SARs.

Feil-Seifer identifies social success as a task oriented benchmark. He asks “does the robot does what it was supposed to do”? Is it being successful in HRIs?

In the social success benchmark we expanded such interpretation and introduced the following categories: type of robotic application delivered and emerging questions, users’ responses, robotic presence, attachment, ethnographic studies and methods used to deliver SARs. The type of robotic application delivered identifies the primary objectives of HRI and emerging questions in such interaction. The users’ responses (e.g. body language, confidence, level of communication and socialization) are extremely relevant to positively/negatively inform researchers about the outcome of HRIs. Robotic presence is a result of how well imitation is delivered (aesthetics) and perceived by elderly groups when using SARs. Attachment deals with the propensity for elderly groups to interact excessively with SARs. To date such levels are unknown but during the robotic workshops we already experienced some signs of attachment associated to the robotic animals’ activities. In ethnographic studies we investigate eventual cultural issues/differences that can inform the content programmed in SARs and map its success. Also the delivering method used by a researcher/performer to deliver entertainment robotic activities with elderly groups is an important step to be considered.

As [6] points out, Roboethics “tries to develop scientific, cultural, technical tools that can assist the development of robots and its diffusion in society”. We think our contribution is part of the new curriculum and practice of Roboethics. In this particular study we conducted robotic workshops with 74 elderly residents (ages 65 plus) in care/extra care homes. Of primary importance is the articulation between the theoretical analysis (ethics) and practical implementation and exercise of SARs. More field work is required and we intend to expand more HRI benchmarks in the course of time. As SARs technologies evolve we will need to revisit the social success HRI benchmark, test it, refine it and evolve our understanding on the emerging ethical challenges of SARs.

REFERENCES


