Using Robots to Modify Demanding or Impolite Behavior of Older People

Heather Draper[1,*] and Tom Sorell[2]

[1] University of Birmingham, Birmingham, UK
h.draper@bham.ac.uk
[2] University of Warwick, Coventry, UK
t.e.sorell@warwick.ac.uk

Abstract. As part of a large scale qualitative study (conducted in France, the UK and the Netherlands) of potential users' views on the ethical values that should govern the design and programming of social robots for older people, we elicited responses to a scenario where a robot is programmed to modify an older person's rude behavior. Participants' responses ranged from outright disagreement with robotized efforts to change characteristic behavior, to approval as a means to an end. We discuss these views against the background of respect for autonomy, the differences and similarities between robot and human carers, and behavior modification in the context of rehabilitation, where the 'no gain without pain' principle is commonly used to justify what would otherwise seem callous. We conclude that such programming may be acceptable in the context of the rehabilitation and promotion of the independence of older people

Keywords: social robots, care-robots, ethics older people, autonomy, behavior rehabilitation, enablement, independence, re-enablement qualitative research, users' views.

1   Introduction

In order to simulate an empathetic response in a care-robot, Patrizia Marti and other FP7 ACCOMPANY (Acceptable robotics COMPanions for AgeiNg Years) researchers at Sienna adapted a Care-O-bot® tablet interface [1]. Sensors in the tablet frame make it touch-sensitive, which enables the user to express urgent need by squeezing the tablet (the 'squeeze-me' facility) [2]. For its part, the tablet is able to display graphic symbols of simulated emotional reactions on the part of the robot. The tablet displays a schematic mask [3] that is easy to read as showing pleasure or happiness and irritation or anger in the context of scenarios developed for user and robot in the ACCOMPANY project. For example, the robot can share the user's supposed happiness at the prospect of a parcel being delivered, and sadness when the user does not drink from a bottle of water (rehydrate) so as to avoid becoming dehydrated when prompted by the robot. An annoyed mask is projected if the user inappropriately uses

---

the squeeze-me facility. Inappropriate use would include always squeezing rather than reserving squeezes for urgent tasks. Part of the rationale for the annoyed mask in this kind of case is to keep the user – who may be quite socially isolated – in touch with social norms of politeness and patience.

As ethicists, we are interested in the ethical implications of robots being used to modify social behavior, and in the reactions of potential users to the idea of a robot that expressed mild annoyance or was assertive with a user, particularly an older user. In this paper, we present and discuss some of the results of a large qualitative study that was designed to enhance the ethics strand of the ACCOMPANY project. More information about the larger study can be found in Draper et al in this volume [4]. The results discussed here pertain to the moral permissibility of temporary refusal by the robot to respond to user commands in the interest of enforcing social norms.[1]

2   Method

We asked 21 focus groups (composed of separate groups of older people, and informal and formal carers of older people) to consider a scenario in which a robot refused to respond to rudely made requests by an older user (see Table 1).

Table 1. Nina Scenario

Nina, who is 70 years old, had a stroke two years ago but has now recovered the use of her arm, though one side of her face droops slightly. She is self-conscious about this, but it does not affect her physical functioning. She is supported at home by a Care-O-bot®. Since having the stroke she has become quite irritable and impatient. She often shouts at her daughter when she visits and complains angrily about her condition. Her daughter finds this very upsetting and has come to dread her visits. Nina has been so rude and demanding that two cleaners have already refused to work for her anymore. She is usually polite with her friends. Her Care-O-bot® has been programmed so that it will not do things for her if she asks sharply or in a demanding tone. It encourages her to say please and thank you and will withdraw help until she does so. Nina finds this infuriating and insists that the Care-O-bot® is reprogrammed to do what she asks no matter how she asks for help.

Our focus groups were convened in three different countries (France, the Netherlands and the UK). Maintien en Autonomie à Domicile des Personnes Agées

---

[1] Carebots programmed to discourage urgent squeezes and with the power temporarily to ignore shouted or rude commands may seem to raise questions about the ethics of persuasive technology [5, 6] but the usual framework for this ethics – the norms of an idealized speech community [5], or a set of criteria for judging samples of persuasive speech and writing [7] – are out of place in the kind of case we consider. We assume that the presence in the home of the carebot is with the user's consent, and also that its various functions, including keeping the user in touch with social norms, are known to the user before they consent to the presence of the robot. Most of the issues raised by this paper concern the voluntariness of the behavior to be modified and the suitability of the robot (as opposed to a human being) as an enforcer of norms governing interactions with people.

(MADoPA) in France convened each kind of group on three separate occasions (n= 9) and Hogeschool Zuyd (ZUYD) in the Netherlands convened two of each (n=6). In the UK the University of Hertfordshire (UH) convened one of each type of group (n=3), and just groups of older people were convened by the University of Birmingham (UB) (n=3). A total of four scenarios were presented for discussion, in the native languages of the participants. This paper concentrates on the scenario that explores improving impolite behavior (Table 1) – a brief outline of the others can be found in Table 2.

Table 2. Brief Description of Remaining Scenarios

Scenario	Brief description
1: Marie	Marie (78) resists the robot's efforts to encourage movement that will help her ulcers to heal. She likes it reminding her to take her antibiotics but not reminders to elevate her leg. She isn't honest with her nurse about how much she is moving.
2. Frank	Frank (89) is socially isolated. His daughter wants him to access an on-line fishing forum with the help of the robot. He isn't keen to try.
4. Louis	Louis (75) likes to play poker online using the robot. He uses its telehealth function to monitor/control his blood pressure. He doesn't let the robot alert his informal carers when he falls (which he does regularly, usually righting himself). His informal carers want to re-program the robot so it will not let him play poker and to alert them when he falls.

The discussions were video or audio-taped and transcribed verbatim. A representative transcript from each type of group (older people, informal carers and formal carers of older people) run in the Netherlands and France was translated into English. All the available English transcriptions were then coded (by Draper) and this coding was independently checked (by Sorell). The results were discussed by the coders and then again with the facilitators at UH, MADoPA and ZUYD until a shared interpretation was reached. The facilitators from MADoPA and ZUYD then coded the outstanding native language transcriptions. Quotations to illustrate the codes were chosen and translated into English and represented in the write up. The report – running into over 70 pages, and containing illustrative quotations – was circulated to all facilitators for verification. A completed data set was also compiled containing all of the coded data.

There were insufficient funds available to translate all of the non-English transcripts. While this will inevitably have affected the reliability of the data, we believe that discussion before, during and after the second round of coding helped to mitigate this limitation. Qualitative methods do not produce quantifiable, generalizable results.

More information about the methodology informing this qualitative study and its analysis can be found in Draper et al [4]. The data is reported using representative quotations that support our interpretation and (where space permits) the spread of the data across individual groups, group types and sites.

3   Results

Similar responses were recorded in all three types of groups. Participants were concerned that Nina's behavior towards her daughter and carers (her 'rude' behavior) could be either a direct result of her stroke, or a response to it. In either case, they felt that it would be difficult to hold her accountable for it, and the correct human response (from her daughter and carers) was therefore tolerance, while the correct robotic response was compliance, however rudely Nina behaved.

> *I can't believe it!* [the programming] *How can it be that people become rude and agitated and everything, when they didn't used to be like that at all? What are you supposed to do if her mind's affected in some way* (pointing to her head)*? You can't tell people like that off!* (MADoPA OPFG1 P3)[2]

> P4: *And it can affect the part of your brain that makes you change your personality.*
> P2: *Yes, it might be that you are physically in pain, or discomfort or something, you know. Yes, exactly... Sometimes it's not that they want to be like that, they can't help it.* (UH IF)

Participants recognized, however, the emotional challenges for the people involved in Nina's care of being tolerant – especially for daughter.

> *I think the daughter definitely needs to ask for help. It's not easy to be sent packing like that.* (MADoPA IFFG1P6)

> *And also if she's rude to her friends they won't come back perhaps ... family will come back no matter how rude you are* (UB OPFG1 P6)

The scenario deliberately left open how responsible Nina might be, as we wanted all of the characters in the scenarios to be both realistic and sympathetic. Participants in all groups were sensitive to the ambiguity.

> *But it is a strange situation because the scenario reads she is nice to her friends. So it is because of her disease, I am friendly or I am not friendly. To the robot she is not nice but for her friends is nice, to them she can talk civilized like "please" and "thank you"   So she is able to do it, so whether it is caused by....* (ZUYD FCFG1 P6)

---

[2] The quotations are coded as follows - <abbreviated name of site> + <type of focus group (OP, IF, FC)> with FG<number of group (1-3) > for sites that held more than one of that type of groups + <participant identifier> to maintain participant confidentiality. Given the space constraints we have only been able to supply indicative quotations. The full data set is held by the corresponding author.

This enabled them to explore the possibility that Nina could, in fact, control her behavior, with consequences for appropriate reactions to Nina on the part of both the robot and human carers. This possibility elicited a range of responses, which we will now briefly outline.

Some participants thought that it was acceptable to program the robot to refuse to cooperate with Nina's rudely delivered requests. There were a range of reasons for this, which were variously combined by participants. Some thought that Nina herself would be better off if she could be a nicer person, as she would enjoy greater continuity of care from her formal carers (who otherwise might refuse to work for her), and that if her daughter (and others) enjoyed visiting, they would come more often. Others thought that rudeness in any form was unacceptable (with some thinking that rudeness even to machines was wrong), whilst some participants thought that her behavior to her carers and daughter was unacceptable and that it was therefore permissible to use the robot to modify it, if possible.

> *I would keep that* [the robot program] *permanently because I don't think that being ill mannered or rude to anybody is the right way for people to live* (UH OP P5)

> *Personally, I think it's really good that the robot doesn't react if she speaks to it too demandingly. I wouldn't like it if someone spoke to me like that.* (MA-DoPA FCFG1 P5)

Unsurprisingly, many participants pointed out that the robot was only "*a machine*" and that it did not therefore matter how rudely Nina spoke to it.

> *Well the carer in terms or the cleaner in terms of their sort of conditions of work and rights that work, right to be respected and to be treated properly by their employer or by anybody else, that's one thing. I think a cleaner has the right to say what you said (*addresses researcher*), y'know, 'Please don't speak to me like that', y'know, 'have respect if you don't mind, or I'm going' But I don't think the robot has… I don't think we can go as far as saying the robot has rights at work* (UBOPFG3 P6)

> *I won't consider it a big problem, if she want to speak in an unfriendly tone, that's fine. The robot won't suffer from it* (ZUYD IFFG2 M5)

A few participants also thought that the fact that the robot was a machine might be a positive advantage. There was some sympathy for the ill or those living with disabilities having an opportunity to vent their feelings, especially on a thing as opposed to a person. A robot, since it lacks feelings and emotions, might be the perfect '*punch bag*' or '*safety valve*' for such feelings.

> *Well, why not because this robot has no feelings so it would be ideal for getting rid of all your aggression ... Because this would be safe, because you do not hurt anyone because it does not feel anything.* (ZUYD FCFG1 P1)

For some participants, however, politeness should extend to all of a person's interactions, even with machines. Here participants seemed to be appealing to personal integrity and control. At other times, rudeness was equated with swearing, and swearing *per se* was disapproved of on that basis.

> *Yes, I don't know. Look, when you start yelling at such a machine it will only get worse…. This is not how you deal with human beings. But it is only a machine. But it is still somebody who helps you.* (ZUYD OPFG2 E3)

Some participants reacted strongly against the idea of robotic attempts to modify behavior – in this scenario and the others. Some of these participants seemed to be appealing to notions of respect for autonomy to justify their reactions, but in at least some cases, their views seem closely linked to the fact that a robot was undermining autonomy. In other words, in other scenarios humans behaving in coercive ways did not provoke the same responses, even though participants were often divided about whether the paternalistic behavior in question was reasonable. Participants who directly appealed to some notion of respect for autonomy were aware that this meant that individuals would have to live with the consequences of their actions.

> *No I don't think a robot should be able to treat somebody as if they're a naughty child... Not not somebody of seventy, no.* (UB OPFGFG1 P6)

> *Personally I'm not sure that the robot should act like that. Basically it's there to help her, she lives with it. If her daughter doesn't like it, she can just visit her mother less often.* (MADoPA OPFG1 P3)

Taken together with the first observation in this section – that participants clearly distinguished between those who had or lacked mental capacity – what emerges is a view about autonomy that goes beyond saying that individuals should be given what they want simply *because* it's what they want. We will be exploring participants' views about autonomy in greater detail in an upcoming ACCOMPANY deliverable, due to be completed by the end of September 2014.

Finally, some participants thought considerations of safety outweighed other concerns in this scenario. Although there were potential benefits to Nina in getting a grip on her rude behavior, the potential risks to her safety from programming the robot to ignore her requests for help outweighed these benefits. Specifically, some of these concerns were based on the robot not being sufficiently sophisticated to be able to distinguish between rude and urgent requests for help.

> *…also it seems that the the the Care-o-bot will not actually do something if she's not polite to it, I think it's dreadful that – [the] machine... actually not do*

what it's supposed to do [4: *frightening*] [2: *I find that quite quite*] *scary. Yeah and I think that's awful to have, to program a machine that that sort of won't help her.* (UoB OPFG2 P5)

Discussion

The reactions of the participants to the range of potential opportunities to change the behaviors of older people presented in the scenarios were interesting. These tended to vary according to participants' perceptions of how usual, beneficial or intrusive the prospective changes might be. We have chosen to discuss the Nina scenario specifically in this short paper, because it lay at the extreme end of a range of behavior-altering interventions, with reminders to take medication at the opposite end and more general health promoting interventions somewhere in the middle.

Responses to Nina were interesting because they tended to focus on Nina's character. The participants seemed to feel that to change Nina's rude behavior was to change her as a *person*, and there was something objectionable about this effort, especially when it came from a robot. Clearly, there is a sense in which our choices help to define us. Their value in this regard can be distinguished from the value of what is chosen. None of the participants thought that Nina's behavior was *acceptable*: they didn't agree with the way she chose to behave towards her daughter and carers. At best, some wanted to say that perhaps her behavior was not the result of something over which she had control. In *this* sense it was not chosen at all, and the correct response – according to them – was therefore for humans to tolerate her, and for the robot to comply regardless of rudeness or inappropriate expressions of urgency.

This too was an interesting result from a philosophical point of view. After all, if she could not control her rudeness, it is not an expression of her autonomy, and respect for her autonomy cannot therefore be used to justify toleration, especially given the apparently harmful effects of her rudeness on others. Her daughter in particular was a captive of Nina's behavior, since, arguably, her filial obligations bound her more tightly than the obligations of Nina's cleaners bound them not to leave her employ.

It is true that carers are supposed to tolerate – or at least regard with some compassionate understanding – grumpy behavior that is provoked by suffering or coming to terms with life-changing conditions. But equally, the sick role requires that patients should co-operate with efforts to assist with recovery [8] in exchange for the suspension of other social norms (like working or being polite). Rehabilitation is meant in part to return a patient as far as possible to the health and independence they enjoyed prior to an adverse event. Against this background, it is reasonable for roboticists to design robots that can help patients like Nina to reconnect with social norms of co-operation by discouraging rudeness. Given that participants did not whole-heartedly agree that human cleaners were wrong to refuse to work for Nina, future research could further explore with potential users why they might object to the robot doing something similar. But we can begin to theorize about this.

Here are two possibilities: (1) The reason why humans but not robots could refuse to suffer Nina's rudeness is that robots are not *able* to suffer from rudeness. They

have no feelings to hurt, no awareness of breaches of social convention that are demeaning to humans and, by analogy, in principle demeaning to themselves. (2) The robot is a thing designed to serve a person, that is, something compliant rather than uncooperative or agenda-setting. It is as if the robot acts out of role when it limits its cooperation or takes the initiative with its mistress. But this line of thought ignores that a carebot is not simply a servant but a servant within a rehabilitative role or a role that maintains the older person's independence with that person's general consent. The norms of rehabilitation rather than the older person's moment-to-moment wishes therefore govern robot-human interactions.

The importance of possibility (2) comes into sharper focus when it is realized that rudeness might not be the only kind of behavior relevant to human-carebot interaction. There is also, more generally, behavior consisting of non-cooperation or indifference to rehabilitation. One of our other scenarios saw a robot trying to encourage an older person (Marie) to move around more in compliance with medical advice. A way of doing this might be to program the robot to limit the number of occasions on which it responded positively to, for instance, getting Marie drinks, on the ground that fetching drinks for herself is a form of therapeutic activity. Here a balance needs to be reached between, on the one hand, ensuring that an older person has the means of rehydration to hand, and, on the other hand, not succumbing to requests that are prompted by laziness, or an unwillingness to suffer some mild discomfort from movement that is beneficial overall. This kind of balance often has to be struck by physiotherapists, for instance, who sometimes have to operate to the principle 'no gain without pain' that can seem callous to the observer.

The operation of the principle 'no gain without pain' is justified in the case of human intervention by the benefits to the patient. The principle operates beyond the area of rehabilitation, since drugs and surgery are often unavoidably accompanied by unpleasant side-effects. Its justification lies not just in the net benefits, but also in the minimization of harmful effects and the agreement of the patient to both the ends and the means. If the application of the principle 'no gain without pain' is justified in these circumstances, then it is justified regardless of whether it is put into effect by a human or a machine, provided that the safeguards are the same.

This suggests that there may be reasons for robots to be programmed not to tolerate what would be regarded as rudeness in human-human interactions in circumstances where rudeness would not be tolerated in human-human interactions. Even in these cases it might be more appropriate for *humans* interacting with difficult patients to assert themselves independently of what the robot does. Arguably the assertion is wrongly delegated to the robot and should properly be undertaken by Nina's daughter and cleaners.

But what if the older person lacked the capacity to agree to both the ends and the means? Would this make a moral difference? Not necessarily. The need for agreement to the ends and means is generated by respect for autonomy; and where capacity is lacking, so too is autonomy. On the other hand, lack of understanding may itself alter the balance of harms and benefits. Understanding why discomfort is necessary can help to diminish its effects. Equally, compassion and understanding that some behavior is not willed can increase the inclination to tolerance and diminish the effects of

what otherwise would be regarded as rude behavior. If such behavior is not willed, it might also not be rude in the strict sense. But there are limits to what humans should be expected to endure even at the hands of those who lack capacity. *This* reasoning cannot be used in the case of robots because they cannot be worn down or stripped of their dignity by being treated harshly, or exhausted by incessant demands.

Conclusion

In this paper we have briefly reported the reactions of focus group participants from France, the UK and the Netherlands to a scenario in which a robot is programmed to modify an older person's rude behavior by refusing to comply with rudely delivered commands. Participants were concerned that Nina's behavior resulted from her stroke, and that the correct response was therefore for her human carers to tolerate it and the robot to comply with her requests regardless of how these were expressed. However, because the scenario was deliberately ambiguous about whether Nina could control her behavior, participants also discussed it as though she was responsible. Reactions varied. Some participants disagreed with the way that the robot had been programmed, because they disapproved of the robot refusing to do Nina's bidding, or because it was a machine that is impervious to rudeness; moreover Nina may benefit from being able to vent her frustrations in a way that did not harm her daughter or carers. Still others thought that the programming disregarded safety considerations. Others took the view that the programming was acceptable (even taking into account that the robot is a machine) because they disapproved of rudeness in general or because the end of improving Nina's behavior justified the means.

In the discussion that followed, we argued that the norms of interaction between care-robots and human beings are not necessarily to be drawn from master-servant relations. Care-robots of the kind being developed in ACCOMPANY, are not primarily at the service of their users, in the sense that their user's wishes are the robots commands. Instead, they engage in routines that help older people to maintain their autonomy in the human world, with the agreement of the older person. There may be a role for the robot as an outlet for unwilled human harshness, and also for persistence in the encouragement of elementary kinds of physiotherapy, again within the context of a rehabilitation or re-enablement plan to which the user consents. The encouragement by the robot of user behavior which conforms to human social norms is more controversial, because it lies at the boundary between what autonomy justifies – being oneself, being nasty and taking the consequences – and the demands of co-operation justified by the goals of rehabilitation or independence. The demands of co-operation are hard to resist reasonably in human-human efforts in the context of rehabilitation; they are not entirely reasonable to resist in the case of human-care-robot interactions – at least when they belong to an agreed plan of rehabilitation or re-enablement. Accordingly, robotic interface designs, such as the ACCOMPANY 'squeeze me' function – that permits the user to summon the robot urgently – can be modified to prevent misuse. Likewise, it is acceptable for the expressive mask to display disapproval of the

user's choices where these undermine efforts towards rehabilitation or the promotion of independence.

References

1. Amirabdollahian, F., Bedaf, S., Bormann, R., Draper, H., Evers, V., Pérez, J.G., Dautenhahn, K.: Assistive technology design and development for acceptable robotics companions for ageing years. Paladyn, Journal of Behavioral Robotics 4(2), 94–112 (2013)
2. Marti, P.: Expression rich communication through a squeezable device, BioRob, São Paulo, Brazil, August 12-15 (2014)
3. Marti, P., et al.: Shaping empathy through perspective taking. In: 2013 IEEE RO-MAN: The 22nd IEEE International Symposium on Robot and Human Interactive Communication, Gyeongju, Korea, August 26-29 (2013)
4. Draper, H., Sorell, T., Bedaf, S., Syrdal, D.S., Gutierrez-Ruiz, C., Duclos, A., Amirabdollahian, F.: Ethical Dimensions of Human-Robot Interactions in the Care of Older People: Insights from Focus Groups Convened in UK, France and the Netherlands. In: Beetz, M., Johnston, B., Williams, M.-A. (eds.) ICSR. LNCS (LNAI), vol. 8755, pp. 138–147. Springer, Heidelberg (2014)
5. Spahn, A.: And Lead Us Not Into Persuasion..? Persuasive Technology and the Ethics of Communication. Science and Engineering Ethics 18, 633–650 (2012)
6. Berdichewsky, D., Neuenschwander, E.: Towards An Ethics of Persuasive Technology. Communications of the ACM 42(5), 51–58 (1999)
7. Baker, S., Martinson, D.L.: The TARES test: Five Principles for Ethical Persuasion. Journal of Mass Media Ethics 16(2), 148–175 (2001)
8. Draper, H., Sorell, T.: Patient Responsibilities in Medical Ethics. Bioethics 16(4), 307–334 (2002)