Assistive Robotics: Research Challenges and Ethics Education Initiatives

[Robótica assistencial: Retos de investigación e iniciativas educativas en roboética]

Carme Torras
Institut de Robòtica i Informàtica Industrial (CSIC-UPC)
c/ Llorens I Artigas 4-6, 08028-Barcelona
torras@iri.upc.edu
http://www.iri.upc.edu/people/torras

Abstract

Assistive robotics is a fast growing field aimed at helping healthcarers in hospitals, rehabilitation centers and nursery homes, as well as empowering people with reduced mobility at home, so that they can autonomously fulfill their daily living activities. The need to function in dynamic human-centered environments poses new research challenges: robotic assistants need to have friendly interfaces, be highly adaptable and customizable, very compliant and intrinsically safe to people, as well as able to handle deformable materials.

Besides technical challenges, assistive robotics raises also ethical defies, which have led to the emergence of a new discipline: Roboethics. Several institutions are developing regulations and standards, and many ethics education initiatives include contents on human-robot interaction and human dignity in assistive situations.

In this paper, the state of the art in assistive robotics is briefly reviewed, and educational materials from a university course on *Ethics in Social Robotics and AI* focusing on the assistive context are presented.

Keywords: Assistive robotics, Machine learning, Roboethics, Ethics education, Science fiction

1. Introduction

The combination of autonomous robots, artificial intelligence and the internet of things offers immense possibilities to improve healthcare and assistance in daily living activities in the coming years. Very ambitious projects are underway in this direction, such as the development of a network where robots can share data and procedures, that is, maps of the buildings visited, acquired manipulation skills, and other learned knowledge, in a common format and independent of the hardware of each one. This network will be connected to the internet of things, where robots can obtain object models and instructions for use for all types of commercial products.

It is often said that these new information technologies represent another step in the social transformation that began with the agrarian and industrial revolutions. Although this is true, a qualitative difference should be pointed out: It is no longer just a matter of automating heavy and repetitive work in the fields and in factories, or that household appliances provide people with free time to enjoy in more creative ways. The difference lies in the fact that these information technologies are designed to *interact with people in their daily environments*, which imposes a series of new technical requirements, as described in the next section, while having some ethical and social implications that will be explained later in Section 3.

The most important requirement of these interactive technologies is the ability to adapt to different environments and situations, as well as to each user (what is known as 'customization'). For this, they need to learn from experiences, i.e., interactions with humans and/or with the environment through sensors and actuators. Adaptability is what allows generalizing from one situation to another, being tolerant of inaccurate perceptions and actions, and developing properly in non-predefined and dynamic environments.

2. Research challenges

The techno-scientific challenges posed by the above-mentioned interactive technologies, as well as the techniques being used to address them, will be illustrated in the framework of some European projects in which my research group is involved. In particular, Socrates [1] project devoted to assist people with mild cognitive deficiencies, and the Clothilde[2] and I-Dress[3] projects, which aim to robotize cloth manipulation in hospital logistics as well as in helping people with reduced mobility to dress.

2.1. Friendly interfaces

For an effective and pleasant person-machine interaction, friendly interfaces are needed. Often these interfaces are multimodal, since they combine text, voice, images and, in the case of robots, also gestures and manipulations. Within the framework of the SOCRATES project, a cognitive training application has been developed to improve the memory of sequences, with three types of interaction (voice, image and robot intervention) that are selected adaptively depending on the situation and the needs of

---

[1] SOCRATES project: http://www.socrates-project.eu/
[2] CLOTHILDE project: https://www.iri.upc.edu/project/show/187
[3] I-DRESS project: https://i-dress-project.eu

the user (Taranović et al., 2018). Similarly, in the I-DRESS project, a multimodal interface has been designed that combines verbal and gestural interaction with the robotic arm, and images of color and depth of the environment, in an application of putting shoes to people with reduced mobility (Jevtić et al., 2019).

Kinesthetic guidance and learning from demonstration has been used to instruct a robotic manipulator to feed a person (Colomé and Torras, 2018b). The scenario used in the food application is shown in Figure 1. By kinesthetic guidance the robotic arm is taught to scoop soup with a spoon from the orange plate or pinch apple pieces with a fork from the blue plate and bring then bring the food to the mannequin's mouth. Since the robot learns the movements in a contextualized manner, in the execution phase it is able to generalize correctly to different positions of the dishes and the mannequin. It is very easy, then, for anyone to teach this ability to the robot.

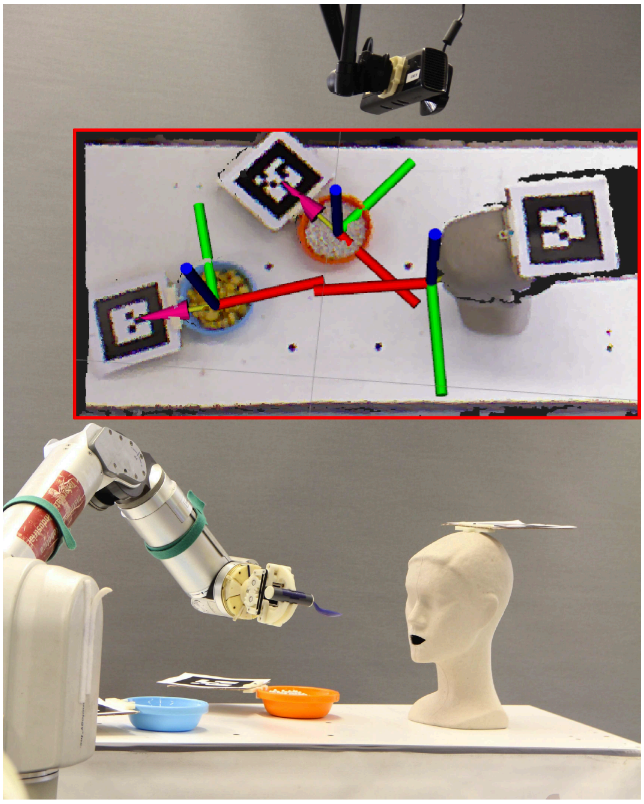

FIGURE 1. Scenario in which the robot is taught to feed a mannequin's head through kinesthetic guidance and learning from demonstration. The red box shows the image captured by the zenithal camera and the location of the dishes and the mouth of the mannequin are displayed. Note that they are marked with QR landmarks to simplify vision requirements.

2.2. Improving through reinforcement learning

Acquiring more elaborate skills (for example, those that involve complex dynamics) only from demonstrations can be very slow or even impossible. Therefore, it has been suggested to use a demonstration to initialize a robot skill, and then explore slight modifications of it in a space of parameters to improve it through reinforcement learning. This is the strategy we have used to teach two robotic arms to fold a polo shirt in the air without letting it go and regrasping it repeatedly by other points (Figure 2). From an approximate demonstration, the robot exercises itself on the task while being monitored by a zenithal camera and evaluates its execution by means of a cost function that measures how well the polo shirt is folded. A reinforcement-learning algorithm is applied that looks for the best policy for folding clothes by the robot (Colomé and Torras, 2018a).

FIGURE 2. Experimental scenario of the task of teaching a bimanual robot to fold garments, in this case a polo shirt, by means of an initial demonstration (left) followed by autonomous learning for reinforcement (right).

2.3. Customization

As mentioned earlier, robot customization is very important in the assistive context to adapt to the abilities and disabilities of the different users and, thus, fully satisfy their needs. It can be attained by building a user's model based on known data, and then refining it through interactions. User preferences such as interests and the way communication has to develop need to be taken into account. Moreover, when the robot has to interact physically with the person, other preferences of physical type come into play, such as distances and interaction speeds, as well as the limitations of movement that the user may have, especially in healthcare settings.

Many elderly people have difficulty putting shoes and they could gain autonomy if a robot would help them do so. But each user has their peculiarities —needs, limitations, tastes— and the robot must be able to adapt to them. Toward this goal, Canal et al. (2019) have proposed a method based on adaptive symbolic planning of the sequence of actions (both movements and communication with the user) to customize the task of putting shoes by means of a robotic arm (Figure 3). With a fuzzy inference system, a user model is built based on the answers given to simple questions and then integrated in the planning domain. The adaptation pursues both task completion and user satisfaction, and it is carried out through a system of variable penalties applied to the rules of the planner. The results show a rapid adaptation of the robot, even when the user's behavior changes or the initial user model is incorrect.

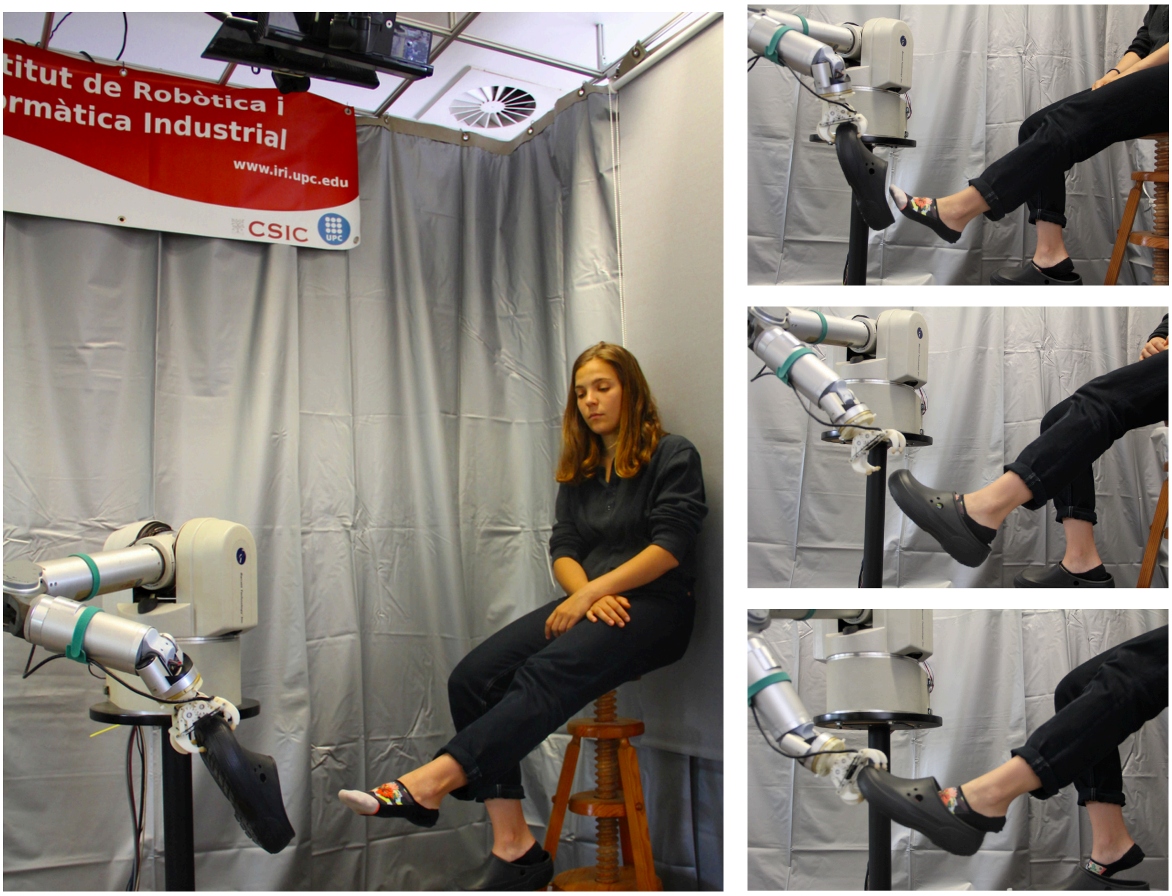

FIGURE 3. Scenario of the application in which the robotic arm puts shoes in a personalized way, after generating and refining a user's model.

Customization can also be carried out at the subsymbolic level to adapt not a sequence of actions, but the way in which one of these actions is executed, whether it is a trajectory, a movement of the clamp, or the actions to take, drag and let go. Canal et al. (2016) developed a method of this type for the task of feeding. Supposing that, in the long run, robots will come from a factory with a repertoire of abilities, the challenge is to ensure that a person without technical training —for example, a caregiver— would be able adapt these abilities to the needs and tastes of a user. Our method is based on learning from demonstration and requires that, during the execution of the task by the robot, the caregiver modifies the position, speed and/or acceleration of the trajectory in the desired direction (Figure 4). The motion and its variance over time are encoded with a probabilistic movement primitive, which has been proven to capture the relevant parameters for a proper execution of the task.

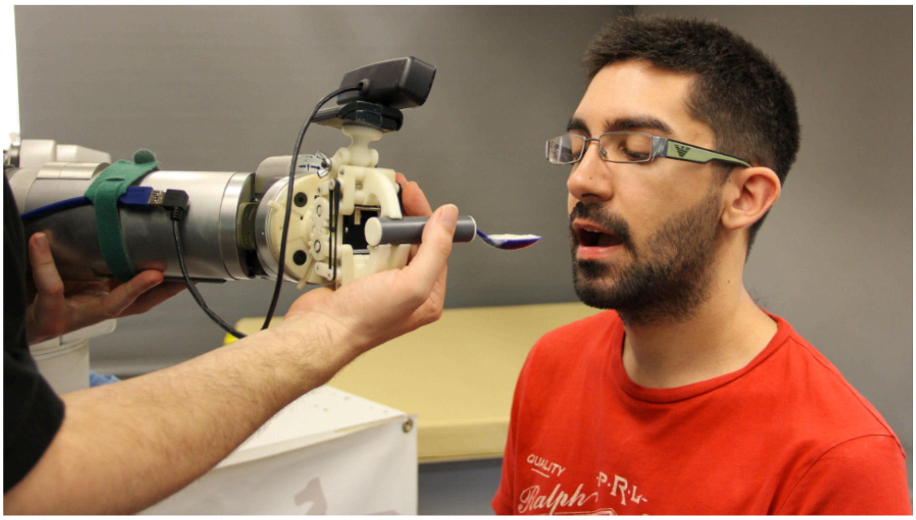

FIGURE 4. A human assistant modifies the way a robotic arm gives food to adapt it to the preferences and limitations of a specific user.

2.4. Safety

Programming a robot to behave compliantly requires a balance between precision and safety, since increasing the precision of the robot (in general, with a high term of error compensation) makes its motion more rigid and, therefore, more dangerous for human beings close to it. This balance is achieved by means of force or impedance controllers based on a model of the inverse dynamics of the robot, which relates the end-effector position, speed and acceleration with the pairs that act on the robot. However, most methods to build a model of this type do not take into account hysteresis in the friction, as in robots such as the Whole Arm Manipulator (WAM) by Barrett Technologies that we have in our laboratory. For this reason, we derived an analytical model of friction for the seven joints of the robot, whose parameters can be adjusted automatically for each robot in particular (Colomé et al., 2015). This allows the robot to easily follow reference trajectories throughout the workspace.

The experimental results show that, using this dynamic model that takes into account friction, the robot is able to learn simple tasks such as putting a scarf, a cap or shoes to a person, safely.

2.5. Handling deformable objects

In human-centered environments, the entities to be perceived are often not rigid objects, but also flexible pieces (e.g., clothes), deformable materials (e.g., food) or even animated beings (e.g., animals or the people themselves). The state of a rigid object is determined by 6 parameters (3 for position and 3 for orientation), whereas the motion of a non-rigid object involves a change in its form, which takes place in a state space of potentially infinite dimension. This huge dimensional leap makes geometric perception techniques developed for rigid objects difficult to apply in this context. For example, in the case of garments, the extension of these techniques requires molding the cloth as a finite element mesh that is deformed by obeying certain restrictions, which leads to an exponential growth of the computation time of any algorithm with the number of nodes of the mesh.

This is why research in this field has focused on the application of automatic learning techniques and, in particular, deep learning has dominated the scene in recent years. In the case of robotics, another option has been to manipulate objects in order to facilitate their perception. This strategy has been the dominant trend in the robotic handling of garments, where clothes are repeatedly regrasped until reaching a configuration that can be easily recognized with simple perception algorithms.

Because manipulating an object to favor its perception is very slow, in our group we have explored the alternative approach of applying complex computer vision algorithms and automatic learning to capture the first piece for the right place to perform the task. For example, from a RGB-D image database of garments with annotated parts (necks, cuffs, waists, hemlines, etc.) as shown on the left of Figure 5, we have developed a method (Ramisa et al., 2014) that, during a training phase, constructs a code using the bag of words learning technique and then uses a support vector machine to classify the parts of clothes according to this code. At run time, a probability distribution is generated from where the wanted part (for example, the neck of a polo shirt) is found, and the best way to take it is determined to achieve a task (for example, place it in a hanger, as shown at the right of Figure 5).

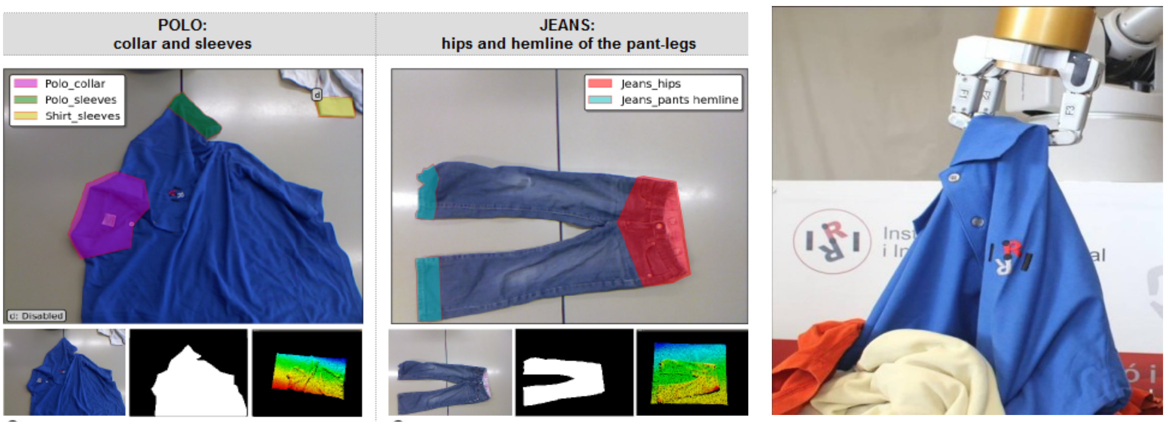

FIGURE 5. Perception and manipulation of clothes. Left: Two elements of the clothing database with the annotated parts (above), as well as the color and depth images

along with the template for background subtraction (bottom). Right: The robot picks up a polo shirt by the neck to hang it.

For a robot to help people to dress —a task we are currently working on in the framework of the European projects CLOTHILDE and I-DRESS—, it is necessary to perceive not only clothing but also the pose of people, where they have their head and limbs, and track them. The most common representation of the human figure is an articulated skeleton (Simó-Serra *et al*., 2017), which for this application must incorporate the volume of the surroundings. We are addressing this volumetric characterization using 3D descriptors (Ramisa et al., 2016), obtained from the point clouds provided by depth cameras (Alenyà et al., 2014).

3. Roboethics

Assistive robots pose a much wider range of ethical issues than their industrial predecessors and other machines, as they enter domains previously exclusive to humans, such as decision-making, feelings, and relationships. The need to regulate their uses for public benefit has led to the establishment of a new discipline: Roboethics, which has come to refer to "Ethics in the age of robots".

The term Roboethics was coined by G. Veruggio at the beginning of the century and refers to the subfield of applied ethics studying both the positive and negative implications of robotics for individuals and society, with a view to inspire the moral design, development and use of so-called intelligent/autonomous robots, and help prevent their misuse against humankind. Subtle distinctions are often made between human ethics applied to robotics, codes of ethics embedded in the robots themselves (sometimes named "machine ethics"), and ethics that would emerge from a potential future consciousness of robots (Veruggio *et al*., 2011). We will here concentrate on the first and touch partially on the second.

There are many ethical theories relevant to robotics. Sullins (2015) briefly surveys consequentialism or utilitarianism (maximizing the number of people that enjoy the highest beneficial outcomes), deontologism (acting only according to maxims that could become universal laws), virtue ethics (relying on the moral character of virtuous individuals), social justice (all human beings deserve to be treated equally and there must be a firm justification in case of mistreatment), common goods (living in a community places constraints on the individual), religious ethics (norms come from a spiritual authority), and information ethics (policies and codes for governing the creation, organization, dissemination, and use of information).

Following a pragmatic option and since no single theory is appropriate for addressing all ethical issues arising in the design and use of robots, we adopt a hybrid approach here. Such hybrid ethics is advocated by Wallach and Allen (2008) as a combination of top-down theories (i.e., those applying rational principles to derive norms) and bottom-up ones (i.e., those inferring general guidelines from specific situations).

Roboethics involves two main areas: legal regulation and ethics education. Regarding the former, institutions such as the European Parliament, the South Korean Robot Ethics Charter, the IEEE Standards Association, and the British Standards Institution are developing regulations for robot designers, programmers, and users. In what follows, we focus on second area, namely roboethics education and dissemination initiatives, particularly at the university level.

3.1. Education initiatives

There are many options to integrate ethics education in technological university degrees, ranging from including a professional ethics course in the syllabus, to allowing students to take some credits or a minor in a Humanities Department, to even offering a mixed degree, like the Computer Science and Philosophy degree at the University of Oxford. Prestigious associations such as IEEE and ACM include 18 knowledge areas in their Computer Science curricula, one of which is "Social Issues and Professional Practice", so that "students develop an understanding of the relevant social, ethical, legal and professional issues". To this end, some courses in this area recur to science fiction stories to exemplify conflictive situations, since narrative is a good way to engage students to safely discuss and reason about difficult and emotionally charged issues without making the discussion personal (Burton et al., 2018). An experience along this line is described next.

3.2. A guide to teach/debate on "Ethics in Social Robotics and AI"

Because of my research on assistive robotics, I became progressively interested in ways to foster a debate and teach Roboethics. This encouraged me to try my hand at fiction, and in the novel *The Vestigial Heart* (Torras, 2018), I imagined how being raised by artificial nannies, learning from robot teachers and sharing work and leisure with humanoids would affect the intellectual, emotional and social habits of future generations. The novel's *leit motiv* is a quotation from the philosopher Robert C. Solomon (1977): «it is the relationships that we have constructed which in turn shape us». He meant human relations with our parents, teachers and friends, but the quotation can be applied to robotic assistants and robot companions, if they are to pervade our lives.

Following a suggestion by MIT Press Editor Marie L. Lee, an appendix with 24 ethics questions and hints for a discussion around the situations appearing in the novel was included in the book, and published together with an online teacher's guide and a 100-slide presentation to deliver a course on *Ethics in Social Robotics and Artificial Intelligence*. It covers six major topics: how to design the «perfect» assistant; the importance of robot appearance and the simulation of emotions for the acceptance of robots; automation in work and educational environments; the dilemma between automatic decision-making and human freedom and dignity; and civil responsibility related to programmed «morals» in robots.

Each section in the teacher's guide follows the same structure, starting with some highlights from the novel, then the corresponding ethics background is provided, followed by four questions and hints for their discussion, and closing with some revisited issues from previous chapters. The section mainly dealing with assistive robots in healthcare is presented below.

3.3. Guide's section on "Human-robot interaction and human dignity"

To contextualize the highlights that follow, here is a summary of the plot: Celia, a-thirteen-year-old girl cryogenically frozen because of her terminal illness, is cured and brought back to life in the 21st Century in order to be adopted. Aside from her memories, she brings something else from the past: feelings and emotions that are no longer in existence. These are what most attract Silvana, a middle-aged woman who works as an emotional masseuse trying to recover the sensations humans have lost. Celia's feelings are also precious research material for Leo, a bioengineer who is designing a creativity prosthesis for the mysterious Doctor Craft, owner of the leading robotics company, CraftER.

3.3.1. Highlights from *The Vestigial Heart*

Chapter 25, pages 177-181:

[Leo:] "Why is it so hard for you people to accept that machines can perform some tasks better than we can?"

[..]

[Silvana:] "What are you trying to do now? Making robots with feelings … and you have to suck them out of a little girl?"

"No, no, please. It's about boosting human creativity by giving people an assistant that stimulates them. [..] It's a device designed by me that helps expand my capabilities. What more could I want?"

"Machines that augment human capabilities seem like a great idea to me: without remote manipulators surgeons couldn't operate on a microscopic scale and, without INFerrers, we'd take too long overthinking the consequences of our decisions … it's ROBs that I reject, and the personal link that is established between them and their PROPs that ends up hogging people's most intimate time and space. You said it yourself: you don't need anything else … and, in the end, you become wooden like them."

"That's what really gets me about the anti-techno lot"—Leo can't take this anymore—"you confuse everything, you get it all mixed up. First off, I was talking about expanding capabilities, not augmenting them. The machines you're so fond of are useful, sure, but they only magnify what we already have. I'm talking about creating new skills, broadening the range of what we can do. For example ROBco …"

[ROBco:] "Question: Would you like a suggestion?" Upon receiving Leo's assent, it goes on. "Try not to repeat yourself. I have already been used as an example and it is obvious that she does not like ROBs. Look for another example, one that appeals to her more."

"Don't you find it degrading when it talks to you like that?"

"Why? It's given me some good advice. Quite the opposite, I'm pleased the prosthesis is working."

Without a doubt this idiot is as wooden inside as he is on the outside. Now he'll make an effort to obey the robot.

<u>*Chapter 28, pages 204-205:*</u>

[Leo:] "What do you take me for? That's the ROB leaving, not me."

[Silvana:] "Of course, I forgot, you built it, so you've already mastered everything it knows how to do."

"Not quite. He accumulates knowledge from lots of different people."

"Okay, okay, I meant that you're not a typical PROP, you take the initiative, not the other way around, like usual."

"I don't understand. All ROBs serve people."

"Exactly. It's just that the service is often poisoned. Why do you think we're against those mechanical contraptions?" She feels she can say this now that the

dummy's not around. "Because we're snobs? Well, no." She's set her course and there's no stopping her now. "Overprotective robots produce spoiled people, slaves produce despots, and entertainers brainwash their own PROPs. And worst of all you people don't care what happens to the rest of us as long as they sell."

3.3.2. Ethical Background and Discussion

Users would expect robot assistants to have the basic interaction competencies to deal with ethically-sensitive situations. This is especially critical in the case of robot caregivers for vulnerable groups, such as children, mentally disabled or elderly people.

Sharkey and Sharkey (2014) identified six major issues to be considered before deploying robot technology in eldercare: (i) opportunities for human social contact could be reduced, and elderly people could be more neglected by society and their families than before; (ii) risk of objectification, if robots would lift or move people around without consulting them; (iii) loss of privacy; (iv) restriction of personal liberty; (v) deception and infantilization that might result from encouraging interaction with robots as if they were companions; and (vi) attribution of responsibility if things went wrong, which opens up the key general concern about the limits of robot decision making in relation to the user's state of mind as addressed under Question A below.

Note that most of these issues are not specific of robots for eldercare, and apply as well to robot companions and even more generally to other types of human-machine interaction… or non-interaction through automatic decision making. This brings us to smart city technologies, such as ambient intelligence and the internet of things, as mentioned in the introduction, which can be very handy in some cases but that, leaving the human out of the control loop, may restrain the freedom and privacy of citizens.

Question A – Could robot decision-making undermine human freedom and dignity?

A feeling of vulnerability similar to that caused by an unforeseen physical contact with a robot may occur at the cognitive level, the solution in this case being much more involved than simply informing the user. Not only is the complexity of the information to be transmitted much higher, but, more importantly, the extent to which a robot should decide and convey its decisions to users depends on their state of mind, which is difficult to evaluate and evolves over time.

Even in the restricted domain of automatic emotion detection—a technology not yet well developed—errors in the interpretation of human mood expressions could strongly impair communication with the user and, more severely, entail danger for the person (e.g., failing to call an emergency service). As Cowie (2015) mentions, the problem is not new, a classical example involving ‘lie detectors’: despite widespread belief in their powers, they were actually much more likely to stigmatize the innocent than to pinpoint the guilty.

Thus, procedures must be devised to ensure that users are not subjected to actions they do not deserve, or not receive responses that they ought to. On a milder scale, provisions should be made for robots to always use respectful language and never intimidate users. In the last highlights above taken from Chapter 25, ***Silvana*** reacts to what she feels is a harsh piece of advice from ***ROBco*** by asking ***Leo*** if he doesn’t find it degrading that the robot talks to him like that.

Boden et al. (2017), in a study carried out under the patronage of the Engineering and Physical Sciences Research Council of the UK, state «a robot used in the care of a vulnerable individual may well be usefully designed to collect information about that person 24/7 and transmit it to hospitals for medical purposes. But the benefit of this must be balanced against that person's right to privacy and to control their own life e.g. refusing treatment.»

A related issue where balance is also needed is whether it is ethically admissible to design robots that can influence human behavior, and if so, whether users must always be aware of robot nudging and how much control they should have over it.

In summary, there is wide consensus that robots and computational systems should be designed in ways that (i) do not denigrate the user to machine-like status, and (ii) do not impersonate human agency by attempting to mimic intentional states leading to deception (Lichocki *et al.*, 2011). Moreover, people should be able to decide whether they wish to interact with these artificial “creatures” and, in case they decide they want to interact only with humans, they should be given the freedom to do so, a guideline that is not easy to implement, as the many companies using chatbots to provide customer support demonstrate.

Question B – Is it acceptable for robots to behave as emotional surrogates? If so, in what cases?

The idea of robot companionship seems natural to some people and almost obscene to others. Levy (2007), in his provocative book and a review of the state of affairs ten years later (Cheok *et al.*, 2017), maintains that many people will no doubt fall in love with robots and that this is completely normal. On the other hand, Bryson (2010) argues that artificial companions should just be servants, machines that you should be able to switch off whenever you like. Sullins (2012) holds an intermediate position in that he accepts people will relate to love machines, and he proposes some ethic design principles to limit the manipulation of human psychology when it comes to building sex robots and simulating love in such machines.

Given the sometimes painful and capricious nature of human relationships, it is not surprising that some might prefer to share their life with a robot, which would have predictable behavior and never criticize, cheat, or disclose their intimacy. This may be acceptable for an adult in full command of their mental faculties, but emotional surrogates should generally be avoided in the case of vulnerable users, and especially children.

The illusion of emotions may have undesired effects on people that are psychologically weak, immature, diminished, or with no technological background, and the risk that they end up being manipulated must be minimized (Boden *et al.*, 2017). Turkle (2007) advises never to disregard that, although the machine may only have simulated emotion, the feelings it elicits are real. Like in other ethical issues discussed up to now, a balance needs to be reached here since, for instance, human caregivers sometimes simulate affection to improve their patient's well-being, and thus robots may also be allowed to do so under similar circumstances.

Let's stress that there is a difference between simulating affection and showing emotional intelligence. The latter entails capturing the emotional state of the user and acting accordingly, which can be very handy in some healthcare situations, but dangerous in the case of interpretation errors as discussed under the preceding Question A.

Robot companionship, even for people with full adult judgment, may have some social consequences as it may lead to sidestep encounters with friends and family, in the end leading humans to no longer privilege authentic emotion, as warned by Turkle (2007). In the case of dependent people there is a symmetrical risk, namely that of allowing friends and family to sidestep their responsibilities. Turkle (2007) touches again on a far-reaching issue when she states, «the question is not whether children will love their robotic pets more than their animal pets, but rather, what loving will come to mean».

The decay of emotions is a recurrent theme throughout the novel. ***Silvana***, an \`emotional masseuse´ that tries to help people recover lost sensations and reads old books to research the power of emotion, sees ***Celia*** as a living example of the feelings that are extinct at the time. Particularly in Chapter 25, ***Silvana*** criticizes that ergonomically-designed technology discourages social relationship, and she strongly argues against robots being built that spoil, corrupt and brainwash people, hogging their most intimate time and space, so that they end up becoming wooden like them.

Question C – Could robots be used as therapists for the mentally disabled?

Some psychologists suggest that the illusion of emotional understanding by a robot that makes eye contact and responds to touch may be therapeutic in some contexts. Additional virtues of robots as therapists are their endless "patience," their capacity for repetitive action without getting "bored," and their never showing unintended feelings, which some humans cannot repress.

Actually, interacting with robots that display social behavior has been shown to help children with autism acquire social skills (Feil-Seifer and Mataric, 2008; Robins *et al.*, 2005). Although the goal of therapy is not to develop an attachment to the robot, it may occur as a side effect and, therefore, the ethical correctness of encouraging such children to engage in affective interactions with machines incapable of emotions is debatable. Whether the finding that severely autistic children prefer featureless, non human-like robots during play (Robins *et al.*, 2004) should be interpreted in favor or against is unclear.

Further to the illusion of emotions discussed above, Turkle (2007) states, «If a person feels understood by an object lacking sentience, that makes eye contact and responds to touch, can that illusion of understanding be therapeutic?» and she continues to ask, «What is the status—therapeutic, moral, and relational—of the simulation of understanding?»

It is worth mentioning that robot-assisted therapy has been applied to other types of patients, such as diabetic children (Lewis *et al.*, 2015; Nalin *et al.*, 2012), with different aims to those for autistic patients: among them, reducing child's stress and anxiety, improving their response to medical treatments, promoting their self-efficacy, and encouraging physical activity. The use of robots in this context raises fewer doubts.

Nonetheless, Riek and Howard (2014) ask, «what happens when the therapy ends and the robot goes away?» Due to possible affective bonds with the robot, its disappearance may have counterproductive effects on the patient, even reversing the benefits of treatment. Thus, these authors suggest that the benefits and risks must be evaluated in advance and protocols must be specified for addressing this circumstance.

Question D – How adaptive/tunable should robots be? Are there limits to human enhancement by robots?

There are two related issues here: up to what extent users should be able to (i) customize robot (possibly, moral) behavior and (ii) enhance themselves by means of robotic prostheses. As regards to the former, it seems clear that, for example, parents should be able to modify the off-the-shelf robot skills to comply with their family values, or caregivers should be able to adapt a robot assistant to the particular needs of a patient. But are there frontiers that such customization cannot trespass? Surely there are, as robots must be prevented from inflicting (physical or psychical) harm to people interacting with them, but setting up the limits is not an easy task.

Turning to the second issue, robotic devices can restore human sensing and physical mobility, thus helping to rebuild body image and restore performance, but they can go beyond that, leading to "human enhancement", i.e., improving human functions beyond what is necessary to sustain and reestablish good health. Again, establishing the limits is tricky: a wearable exoskeleton connected to the spinal cord of a stroke patient may restore their walking ability, and artificial retinas may palliate visual deficiencies, but it is not hard to imagine other uses of bio-robotic prostheses that may turn a human into a cyborg or a living weapon, maybe even remotely controlled by someone else. This extends to cognitive enhancement as well. One of the main themes of the novel is ***Dr. Craft***'s determination to get (and keep only for himself) a "creativity prosthesis" that enhances his inventive capacity, and ***Leo*** is in charge of developing it.

The debate is ultimately polarized into two main positions: transhumanists and bioconservatives. Transhumanism holds that the current form of the human species, on both somatic and cognitive levels, is merely a specific stage of human evolution, and we have only begun to grasp the extent of possible future integrations between the natural and artificial. Bioconservatism stresses the need to investigate the significance and the implications of the transformations concealed behind the apparently neutral technological development involving humans, thus placing the concepts of nature and human dignity as insurmountable limits (Palmerini *et al.*, 2016).

The challenge is how to ensure that robots improve the quality of our daily lives, widen our capabilities, and increase our freedom, while avoiding their making us more dependent and emotionally weak; that is, the eternal dilemma of how to take the good without suffering from the bad side-effects. In their heated discussions, ***Leo*** defends the positive view of robots as enhancers of our physical and cognitive capabilities, while ***Silvana*** highlights the downside that relating to robots ends up replacing people's intimate relationships.

4. Conclusion

Our foreseen growing interaction with robot assistants and all sorts of devices in everyday life poses important research challenges —both technical and scientific, as well as in the humanities and social sciences— with a lot of potential to substantially shape the future and which are fostering an interesting social and ethical debate.

Philosophy, psychology, and law are providing perspectives and prior knowledge to the debate, while science fiction permits freely speculating upon potential scenarios and the role that humans and machines will play in the *pas de deux* that irredeemably connects us. Along this line, educational materials based on science-fiction stories have proven very effective in engaging students taking roboethics courses (Burton *et al*., 2018).

Acknowledgement

This work is partly supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme through the project CLOTHILDE - CLOTH manIpulation Learning from DEmonstrations (Advanced Grant agreement No 741930), and the Spanish Research Agency through the María de Maeztu Seal of Excellence to IRI (MDM-2016-0656).

References

Alenyà G., Foix S., Torras C. (2014) Using ToF and RGBD cameras for 3D robot perception and manipulation in human environments. *Intelligent Service Robotics*, 7(4): 211-220.

Boden M., Bryson J., Caldwell D., Dautenhahn K., Edwards L., Kember S., Newman P., Parry V., Pegman G., Rodden T., Sorrell T., Wallis M., Whitby B. and Winfield A.F. (2017) Principles of robotics: Regulating robots in the real world. *Connection Science*, 29(2): 124-129.

Bryson J.J. (2010) Robots should be slaves. In Wilks Y. (ed.): *Close engagements with artificial companions: key social, psychological, ethical and design issues*, pp. 63-74, Amsterdam, The Netherlands: John Benjamins Publishing Company.

Burton E., Goldsmith J., Mattei N. (2018) How to teach computer ethics through science fiction. *Communications of the ACM*, 61(8), 54-64.

Canal G., Alenyà G., Torras C. (2016) Personalization framework for adaptive robotic feeding assistance». *8th International Conference on Social Robotics*, Lecture Notes in Artificial Intelligence, vol. 9979, pp. 22-31, Springer.

Canal G., Alenyà G., Torras C. (2019). Adapting robot task planning to user preferences: An assistive shoe dressing example. *Autonomous Robots*, in press.

Cheok A.D., Levy D., Karunanayaka K., Morisawa Y. (2017) Love and Sex with Robots. *Handbook of Digital Games and Entertainment Technologies*, Springer Singapore, pp. 833-858.

Colomé A., Planells A., Torras C. (2015) A friction-model-based framework for reinforcement learning of robotic tasks in non-rigid environments. *IEEE Intl. Conf. on Robotics and Automation (ICRA),* Seattle, USA, pp. 5649-5654.

Colomé A., Torras C. (2018a) Dimensionality reduction for dynamic movement primitives and application to bimanual manipulation of clothes. *IEEE Transactions on Robotics*, 34 (3): 602-615.

Colomé A., Torras C. (2018b) Dimensionality Reduction in Learning Gaussian Mixture Models of Movement Primitives for Contextualized Action Selection and Adaptation. *IEEE Robotics and Automation Letters*, 3 (4): 3922-3929.

Cowie R. (2015) Ethical issues in affective computing. The Oxford handbook of affective computing, 334.

Feil-Seifer D., Mataric M. (2008) Robot-assisted therapy for children with autism spectrum disorders. Proc. 7th Intl. Conference on Interaction Design and Children, pp. 49–52.

Jevtić A., Flores A., Alenyà G., Chance G., Caleb-Solly P., Dogramadzi S., Torras C. (2019) Personalized robot assistant for support in dressing. *IEEE Transactions on Cognitive and Developmental Systems*, in press.

Levy D. (2007) *Love, Sex with Robots: The Evolution of Human-Robot Relationships*, Harper Collins Publishing: New York.

Lewis M., Oleari E., Pozzi C., Cañamero L. (2015) An Embodied AI Approach to Individual Differences: Supporting Self-Efficacy in Diabetic Children with an Autonomous Robot. Tapus A., André E., Martin J-C., Ferland F., Ammi M. (eds.): *Social Robotics*. Lecture Notes in Computer Science, 9388, pp. 401-410, Springer.

Lichocki P., Kahn Jr P.H., Billard A. (2011) A survey of the robotics ethical landscape. *IEEE Robotics and Automation Magazine*, 18(1), 39-50.

Nalin M., Baroni I., Sanna A., Pozzi C. (2012) Robotic companion for diabetic children: emotional and educational support to diabetic children, through an interactive robot. Proc. 11th ACM Intl. Conference on Interaction Design and Children (IDC'12), New York, pp. 260–263.

Palmerini E., Azzarri F., Battaglia F., Bertolini A., Carnevale A., Carpaneto J., Cavallo F., Carlo A.D., Cempini M., Controzzi M., Koops B.J., Lucivero F., Mukerji N., Nocco L., Pirni A., Shah H., Salvini P., Schellekens M., Warwick K. (2016) Robolaw: Guidelines on Regulating Robotics.

Ramisa A., Alenyà G., Moreno-Noguer F., Torras C. (2014) Learning RGB-D descriptors of garment parts for informed robot grasping. *Engineering Applications of Artificial Intelligence*, 35: 246-258.

Ramisa A., Alenyà G., Moreno-Noguer F., Torras C. (2016) A 3D descriptor to detect task-oriented grasping points in clothing. *Pattern Recognition*, 60: 936-948.

Riek L., Howard D. (2014) A Code of Ethics for the Human-Robot Interaction Profession. *Proc. We Robot*, pp. 1-10.

Robins B., Dautenhahn K., Te Boerkhorst R., Billard A. (2004) Robots as assistive technology - does appearance matter? *13th IEEE Intl. Workshop on Robot and Human Interactive Communication (RO-MAN)*, pp. 277-282.

Robins B., Dautenhahn K., Te Boerkhorst R., Billard A. (2005) Robotic assistants in therapy and education of children with autism: Can a small humanoid robot help encourage social interaction skills? *Universal Access in the Information Society*, 4(2), 105–120.

Sharkey A., Sharkey N. (2012) Granny and the robots: ethical issues in robot care for the elderly. *Ethics and information technology*, 14(1), 27-40.

Simo-Serra E., Torras C., Moreno-Noguer F. (2017) 3D human pose tracking priors using geodesic mixture models. *International Journal of Computer Vision*, 122 (2): 388-408.

Solomon R.C. (1977) *The Passions*. New York: Anchor Press / Doubleday.

Sullins J.P. (2012) Robots, love, and sex: the ethics of building a love machine. *IEEE Trans. on Affective Comput*ing, 3(4), 398–409.

Sullins J.P. (2015) Applied professional ethics for the reluctant roboticist. In *The Emerging Policy and Ethics of Human-Robot Interaction*, edited by L.D. Riek, W. Hartzog, D. Howard, A. Moon and R. Calo, Workshop at the 10th ACM/IEEE International Conference on Human-Robot Interaction, Portland.

Taranović A., Jevtić A., Torras C. (2018) Adaptive modality selection algorithm in robot-assisted cognitive training, 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, pp. 4456-4461.

Torras C. (2018) *The Vestigial Heart*. *A Novel of the Robot Age.* MIT Press. Cambridge, Massachusetts. http://mitpress.mit.edu/books/vestigial-heart

Turkle S. (2007) Authenticity in the age of digital companions. *Interaction Studies*, 8(3): 501-517.

Veruggio G., Solis, J. and Van der Loos M. (2011) Roboethics: Ethics applied to robotics [from the guest editors]. *IEEE Robotics and Automation Magazine*, 18(1): 21-22.

Wallach W. and Allen C. (2008) *Moral machines: Teaching robots right from wrong.* Oxford University Press.