Human Arenas
https://doi.org/10.1007/s42087-020-00125-7

ARENA OF HEALTH

First, They Came for the Old and Demented:

Care and Relations in the Age of Artificial Intelligence and Social Robots

Henrik Skaug Sætra[1]

Received: 4 April 2020 / Revised: 4 June 2020 / Accepted: 16 June 2020
Published online: 04 July 2020

Abstract
Health care technology is all the rage, and artificial intelligence (AI) has long since made its inroads into the previously human-dominated domain of *care*. AI is used in diagnostics, but also in therapy and assistance, sometimes in the form of social robots with fur, eyes and programmed emotions. Patient welfare, working conditions for the caretakers and cost-efficiency are routinely said to be improved by employing new technologies. The old with dementia might be provided with a robot seal, or a humanoid companion robot, and if these companions increase the happiness of the patients, why should we not venture down this road? Come to think of it, when we have these machines, why not use them as tutors in our schools and caretakers for our children? More happiness reported, as our children are entertained, well-nourished, well-trained and never alone. Lovely and loving robots have also been made, and happiness abounds when these are provided to lonely adults. Happiness all around, and a hedonistic heaven – the utilitarian's dream, as reported, or measured, well-being reaches all-time highs. But there is a reason to be wary of this development. The logic that allows this development ultimately leads to the conclusion that we would all be best off if we could simply be wired to a computer that provided us with whatever we needed to feel perfectly satisfied. The care-giving machines are here.

Keywords Care · AI · Social robots · Deception · Hedonism · Ethics · Utilitarianism

Introduction

Health care technology is all the rage, and artificial intelligence (AI) has long ago made its inroads into the domain of care. AI is used in diagnostics, but also in therapy and assistance.

✉ Henrik Skaug Sætra
Henrik.satra@hiof.no

[1] Østfold University College, Remmen, 1757 Halden, Norway

Some of the AIs are robots—embodied artificial intelligence—in human, animal, or another form. These robots are more than cold steel, however. They often have fur, eyes and make-believe emotions.

The benefits of the various ways in which we use AI is great, and patient welfare, working conditions for the health care worker, and cost-efficiency are routinely said to be improved by these developments. The old with dementia might be provided with a robot seal, or a humanoid companion robot, and if these companions increase the happiness of the patients, why should we not venture down this road?

Come to think of it, when we have these machines, why not use them as tutors in our schools and caretakers for our children? The social robot is home, waiting with perfect and nutritional food, endless attention, and intelligence in the realm of homework that surpasses that of most parents. More happiness is reported, as our children are entertained, well-nourished, well-trained and educated, and never lonely. We also know that some adults do not have partners. Lovely and loving robots have also been made, and happiness abounds when these are provided to the lonely. Happiness all around, and a hedonistic heaven – the utilitarian's dream, as reported, or measured, well-being reaches all-time highs.

But something changes as we replace humans with machines in the areas of life where human relations are most important. I argue that replacing an industry worker with a robot is not the same as replacing a caretaker with a robot, and that the logic that allows this development ultimately leads to the conclusion that we would all be best off if we could simply be wired to a machine that provided us with whatever we needed to feel perfectly satisfied.

The machines of well-being are here. First, they came for the old and demented. Then they came for our kids. In the not too distant future, a *lot* of us will be both old and demented, but the social robots might be coming for us even before we get that far. If we do not clearly state what relationships we *want* to outsource to robots, there is little to suggest that robots or other technologies will not invade and supplant humans in *most* relationships.

In this article I argue that we should be wary of these developments. While social robots show great promise in certain respects, there are certain drawbacks associated with people's rights, autonomy, liberty, and, in the end, *dignity*, as deception and infantilization are issues inherent in the debate about robot in care. A decision must be made regarding what it means to have a good life and what well-being is. Burr et al. (2019) discuss the ethics of 'digital well-being', and 'the impact of digital technologies on what it means to live a life that is good for a human being'. I do the same, as I consider how the use of social robots influences our well-being. This article discusses the ethics of social robots and how we can take a principled approach to *when* and *how* we employ them. I focus on how *ethical hedonism*, *consequentialism* and *utilitarianism*, where the maximization of pleasure and pain is taken to be the basis for evaluating right and wrong, will lead us astray. In doing so, I discuss the issue at hand in the terminology of political theory, responding to Coeckelbergh's (2018) call for the use of political philosophy in order to better understand the challenges created by new technologies. In particular, the challenge of trying to understand where we will end up if hedonism and consequentialism are allowed to guide our employment of technology.

What Is AI and Social Robots?

Artificial intelligence (AI) is, in simple terms, *non-natural intelligence*. Hobbes distinguishes natural and artificial by stating that *nature* is what God has created, while the artificial is our attempt to imitate God's work (Hobbes 1946). God or no God, such an understanding will suffice, as artificial intelligence is fundamentally understood as humanity's attempt to replicate parts of their own mental capabilities in machines. I adopt a broad approach to the definition of artificial intelligence and will consider systems that can perform tasks we usually believe require intelligence to be *artificially intelligent* (Brundage et al. 2018).

Turing wrote *Computing Machinery and Intelligence* in 1950. There, he described a test to determine whether or not a machine is intelligent. In order to pass, a machine must be able to fool a human into thinking that it is not a machine (Turing 2009). As noted, my demand for *intelligence* is far less strict. For my current purpose, I am interested in the *idea* of the Turing test for two reasons. Firstly, some people will believe that a social robot *is* in fact alive, in which case some version of the Turing test is clearly passed. Secondly, a machine can be life-like enough to have interesting effects on our behaviour even if it is *not* capable of rationally persuading us that it is not a machine. I return to this in Section 5.3. I will not deal with the intricacies of the various forms of the Turing test, as this simplified understanding of the concept suffices.

I discuss *social robots*, and *the robotic moment* is Turkle's (2011) phrase for describing our acceptance of these robots. The definition of a *robot* is harder to pin down than one might at first imagine. As Gunkel (2018) shows, it is a moving target. Winfield (2012) defines a robot as an artificial device that both senses, and purposefully acts in, its environment. He also points to other definitions which emphasize (a) the embodiment of AI or (b) a machine's ability to autonomously do 'useful work'. For my current purpose, such a definition will suffice. In our relations with them, we rarely seem to bother about what these intelligences really *understand*, or know (Turkle 2011). Latikka et al. (2019) provide an up-to-date account of acceptance of robots and show, amongst other things, that young people are more accepting than older people, people with much experience with technology are more accepting than those with little experience, and men more so than women. However, it is suggested that robots are less whole heartedly welcomed into the care sector than in the industrial sector and there are differences between social and non-social robots (Savela et al. 2018; Latikka et al. 2019).

First, They Came for the Old and Demented...

Robots as Companions and Therapists

Once, an economist told us that in the long run, we will all be dead (Keynes 1923). Today, scientists of sciences less dismal are telling us that in the long run, a whole lot of us are going to have dementia. In just 11 years, 57.62 million might suffer from dementia and 135.46 million in 2050 (Prince et al. 2013). Mordoch et al. (2013) cite the World Health Organization which predicts that the number will be 115.4 million people in 2050. As more and more people get dementia, more and more compassionate caretakers are required (Mordoch et al. 2013). It is these caretakers I focus on in this article—not dementia in itself, or its causes.

No wonder, then, that we are looking for ways to deal with this troubling wave of old and demented people. We must prepare and fortify. These days, one of the most popular lines of defence is the use of robots in care (Poulsen and Burmeister 2019). Dementia is not the only challenge, as demographic change *in general* will pose a challenge to our societies. However, dementia is the problem which is most often mentioned in relation to old age and associated medical conditions (Bodenhagen et al. 2019).

One aspect in which AI shows great promise is in the diagnosis of dementia. A study shows that a deep learning model could predict an Alzheimer diagnosis long before (on average *6 years before*) the final diagnosis was actually made (Ding et al. 2018). I focus, however, on the use of *social* robots. But robots have also been developed for all sorts of manual tasks involved in eldercare, such as feeding, lifting and washing (Sharkey and Sharkey 2012a). In their description of *The Eldercare Factory*, Sharkey and Sharkey (2012a) describe how robots are now also used for *companionship*.

One unlikely door through which robots entered the world of dementia was through the use of *animals* in therapy sessions—animal-assisted therapy (AAT) (Downes et al. 2013; Roger et al. 2012). Animals have been shown to calm patients and improve both their moods and social interactions (Bernabei et al. 2013). Behaviour and mood might, however, worsen if animal therapy is subsequently withheld (Soler et al. 2015). Animal presence elicits smiles and visual contact and even increases verbalisation (Soler et al. 2015).

But clever minds are never at rest, and someone soon realized that the furry little things we make as shells for our cold and hard robotic mechanical cores carry a great resemblance to animals. Furthermore, while biotic animals carry the heft of ethical issues, smell and faeces, death and illness, tempers and even wills of their own, our abiotic animal therapists have no such issues. Real animals might also be violent and dangerous (Soler et al. 2015). Furthermore, patients may be violent, too, and the ethical aspects of placing animals in potentially dangerous situations are plentiful and complicated (Soler et al. 2015). Allergies and costs can also make AAT difficult in certain situations (Soler et al. 2015).

The robots have perfect round eyes—larger and bluer than any animals plagued by biology. Perfect tempers all around. And, the icing on the cake: they are quite smart, or at least as smart as we want them to be. We can program their behaviour so that they encourage the behaviour we want to see in the biotic patients—the old and demented. Robots suffer no old age and no sickness. Just plug it in, recharge, and you are good to go. The *real* is cumbersome, so *artificial* animals—robots—are introduced (Soler et al. 2015).

Two reasons are often cited in favour of robots in care of the elderly: *rationalization* and *improved quality of care* (Bemelmans et al. 2012). In a study of the effects of companion robots in care for the elderly, Bemelmans et al. (2012) found positive effects on psychological factors, such as mood, loneliness, social connections and communication and physiological factors, such as stress reduction. The effects of robots are often given in comparison either with no intervention or real animals, but the impact compared with *human interaction* is often not mentioned. In such a comparison, Sharkey and Sharkey (2012a) say, the robots might not fare so well.

Mordoch et al. (2013) use the term *social commitment robots* in their literature review of the use of such robots in the care of the elderly with dementia. These robots are 'designed to promote therapeutic interactions through communication and social interaction' (Mordoch et al. 2013). These are *interactive* robots which provide personal 'interactions, pleasure and relaxation' (Mordoch et al. 2013). The name *therapeutic* thus relates to the therapeutic effects these robots provide—increased well-being—which are in most respects similar to the effects of AAT as discussed above.

One of the most famous robot companions is *Paro*, the robotic seal (Paro Robots 2018; Wada et al. 2008). It is branded as a therapeutic robot and is used both to relieve stress and loneliness in the elderly as it is employed in elder care facilities (Paro Robots 2018). It is said to provide the documented benefits of *animal* therapy, even though it is no animal per se (Paro Robots 2018). It is somewhat smart, as it can remember the name it is given, and it will also learn and adapt its behaviour in order to encourage interaction (Paro Robots 2018). In a study by Wada et al. (2008), they used Paro in therapy for patients with dementia, and they state that it has a 'high potential to improve the condition of brain activity in patients suffering from dementia'. Of interest here is, as mentioned, not a comparison to *nothing* but a comparison to having interactions with other people.

Another application of AI and robotics to the group of people who suffer from dementia is in the form of *personal assistants*. There is good reason to believe that many could benefit from having robots designed to provide important cues for action that might otherwise be forgotten, providing safety in terms of checking stoves, ovens, etc. (Sharkey and Sharkey 2012a). Some will most likely also prefer robot assistance for, say, bathing, going to the bathroom, etc. (Sharkey and Sharkey 2012b). Assistive technology (AT) can help residents 'age-in-place' and as such provides great potential benefits (Vandemeulebroucke et al. 2019). I focus, however, mainly on the *social* robots employed in care.

If we were to consider the arguments in favour of social robots in care, without considering costs, it seems to add up to the idea that any device that improves a patient's mental and physical well-being should be introduced, regardless of biotic status. This will serve as the first proposition in the argument in favour of social robots:

> A therapeutic device that increases a patient's mental and physical well-being should be employed.

Health Care and Efficiency

Realism is an approach to politics that encourages us to look at how things *really* are, instead of how we wish they were. Machiavelli (2003) and Hobbes (1946) provide prime examples of this approach. In health care and practical politics, we quickly get into the realm of realism, as the *ideal* solutions are never really feasible. The best available medicine is too expensive to be distributed to all that could benefit from it, the best treatment often carries prohibitive costs, and no society can afford to assign a personal nurse to all the people in need of assistance.

When economics matter (which, according to some, would be *always*), it is tempting to substitute human beings with technology: particularly in the case of the elderly, as demographic projections warn of a future with *more* old people and a *shrinking* workforce. This creates a double bind, and our societies will have to face it in *some* way. Perhaps by the use of robots, as robots have the potential to make the care sector much more effective in terms of getting things done without human beings.

I will not discuss this aspect in detail but will take it as a given that costs need to be considered when determining what level, and kind, of care a society is to provide. I will not examine the ideal but focus on the *real* world of politics and care. These considerations provide the second proposition for the argument in favour of social robots:

> When all else is equal, one should arrange society so as to improve the well-being of as many people as possible.

However, even a realist will consider ethics, and the foregoing proposition has an ethical underpinning: that of consequentialism. Utilitarianism is perhaps the most famous version of consequentialist ethics, and it is often relayed as the quest for maximum *happiness*—or *well-being*—which is the word I use in these propositions (Mill and Bentham 1987). This discussion is continued below.

When Technology Replaces Human Contact

Mordoch et al. (2013) remind us that 'human contact is a critical component of care' and this means that there are several important ethical aspects to consider with regard to social robots. First of all, robots might replace human contact (Chiberska 2018; Sharkey and Sharkey (2012a). Mordoch et al. (2013) say that robots should 'augment and bridge' human contact, but we could easily see futures in which those in power decide that robots are so much cheaper to deploy than humans that we *replace* humans instead of *augmenting* them. Chiberska (2018) mirrors this notion, as she says that social robots 'can be used to enhance person-centred care rather than to replace human contact'. While we need innovation, we must be wary of 'undue infantilization and deceptions and tendencies to overly reduce human contact' (Mordoch et al. 2013). My assumption is that people of old age have both *rights* and *dignity*, even when demented. Ethical concerns must be emphasized, and ethical guidelines developed (Mordoch et al. 2013). Sharkey and Sharkey (2012a) also mention that people *need* (human) companionship for the preservation of their mental health.

When Robots Become Our Companions

I do not discuss *all* current use of AI, but our use of AI as *social companions*. This includes all sorts of relations with machines that include some form of relationship that is more than purely rational or instrumental (from the perspective of the parties in the relation). While our introduction of a robot seal in an eldercare facility might have a purely instrumental, and rational, basis on the part of the managers, or politicians, the patients develop *social* relations with the seal. They will respond emotionally to the robot. This will be the case even for those that have a rational understanding of the fact that the seal is an abiotic entity. I return to this in the discussion of deception below.

We currently interact with machines in ways that were previously reserved for relations between people and their fellow biotic beings. We even make robots that replace other humans for the function of *love*, as Levy (2008) discusses in *Love and Sex with Robots.* Roxxxy, the loving robot of the somewhat superficial kind, carries conversations, and brothels with robot staff are no longer (just) a fantasy (Scheutz and Arnold 2016; Lockett 2017). *RealDollX* is a companion, 'made to fall in love' (Realbotix 2019). Bendel (2014) notes that companion robots are more important than some seem to believe, as sexual health is important for both general health and well-being. In an analysis of media representations of love with robots, Döring and Poeschl (2019) note that non-fictional accounts of robot love tend to focus on sex, while the *fictional* accounts far more often problematize and describe the deeper emotional aspects of intimate relationships. Roxxxy explicitly aims for *intimate* relations, and this raises certain questions, as robots have not been party to such relations in any significant degree up until now.

However, as I have shown, we *have* used robots for companionship for a long time, and I will argue that the difference between these relationships is more a difference of *degree* than a difference of *kind*. The main reason we have not used robots to fulfil more of our relational needs to date might simply be that more intimate relationships are more demanding. Until now, we have not had machines that could satisfy these demands to a sufficient degree.

Just as the artificial pets had plenty of advantages over their live counterparts, robot companions in general have many advantages over other human beings. They have unlimited patience, attention, fidelity, endurance, etc. (Levy 2008). Some, however, question the wisdom of uncritically accepting this development.

Robots Are *Live Enough*

As Turkle (2011) has convincingly shown, primitive robots are quite easily equipped with features that make them *live enough* for us to respond to them *as if* they were alive. One problem with certain robots is that people mistreat them. They trick them, play with them and even abuse them. In an attempt to stop this, a company in Finland equipped a robot with a pair of googly eyes. Once the eyes were in place, the abuse stopped, and the reactions to the robot became more positive (Schwab 2019).

In the domain of animation, the effects of subtle things such as this are well known. Johnston and Thomas (1995) published their classic book *The Illusion of Life: Disney Animation*, in 1981, and *exaggeration* is one of the 12 basic principles of animation. While large and exaggerated eyes are particularly effective in eliciting an emotional response, *any* set of eyes might actually do. Bateson et al. (2006) have shown that simply having a poster depicting a pair of eyes triggers certain social cues in us—causing those they studied to give *three times more* to a shared coffee fund than the subjects in a non-poster-with-eyes condition.

We may pride ourselves as a species for being both rational, smart and quite sophisticated. However, we are more easily deceived than we might like to acknowledge. Life imitating features in robots *cue* certain responses in us.

The creators of Paro have clearly attempted to imitate life. Successfully so, as it manages to fool people into believing that it is a real seal. Its creators observed real seals while developing Paro, and they have programmed proactive behaviour and responsivity into it, with the express goal of encouraging interaction and the feeling of a real connection with its often unsuspecting users (Wada et al. 2008). These examples just serve to show that even very simple adjustments to machines and robots have clear effects on our behaviour. It is important to note that these effects are due to their appeal to subconscious and prerational mechanisms. As discussed in detail in Sætra (2020), social AI is in a certain sense *parasitic* to human social mechanisms.

What Happens to Us?

Turkle (2011) has studied what happens to human beings when our relationships increasingly consist of relationships with machines. Human relationships consist of intimacy and authenticity, she says, and she argues that this cannot be the case for relationships with machines (Turkle 2011, p. 6). Damasio (2018) agrees and states that robots can have neither life nor feelings and the prospects of them getting such things are slim to none.

The question of *authenticity* is important, as how we approach this question has deep consequences for how we view both ourselves and the potential of machines. What are *authentic* feelings, for example? If we take the view that they are nothing but chemical reactions, the door for authentic *machine feelings* is wide open.

If we believe that human beings are *more* than what can be simulated by sophisticated machines, interacting heavily with machines that fool us into thinking that they *are* authentic might *diminish* us (Turkle 2011). Authentic relationships with human beings are characterized by surprises, joys, disappointments, astonishment, comfort and all such things. Warts and all – authentic relationships are not valuable because of their intimation of *perfection*, but because of their varied and *real* character. Turkle (2011) mentions that our relationships are formed and shaped by 'history, biology, trauma and joy' and this, she says, robots cannot aspire to. Metzler et al. (2016) disagree and state that robots *might* become authentic companions. This is supported by the developments discussed by Cominelli et al. (2018), as they discuss how robots are now built with intimations of *emotions*.

Damasio (1994, 2003) discusses the role of emotions in what we understand as *reason*, and in *The Strange Order of Things* (2018), he explains why he believes that machines will *not* achieve perfect imitation of the way human beings function in this respect.

Computers behave *as if* they have feelings, and the tragedy, if we believe that our relations with computers might be harmful, is that these *as-if* performances are effective enough for us to perceive them as real. A central point is that even if we do not believe the robots to be alive, sentient, etc., on a conscious level, we still seem to consider them 'alive enough' to respond emotionally to them and bond with them (Turkle 2011). This is the case even for robots that are *not* designed for this purpose. A therapeutic seal, or a toy, may be designed simply to be cute and fun, but these intentions are of little consequence if we respond to these machines on a much more fundamental level—we construe the relations with machines as *intimate* (Turkle 2011). Their empty artificial eyes look at us as a result of computer code and programming, but we perceive something alive, something with feelings, something in need of nurture and care. So, we bond. We start to *care* for machines; we become reluctant to part with them, and grieve when they break (Turkle 2011). Even very dumb machines can be *alive enough* for us to respond this way.

This, you might argue, is little different from the situation which occurs when a child conjures up life in their stuffed toys and bonds with them (NOU 2011:11 2011). Turkle (2011), however, points out that this *is* a very different thing, as children projecting themselves upon such objects are different from us treating these new machines as *subjects*—alive and truly separate from us.

Perhaps the most important question is one posed by Turkle (2011), as she asks: 'What if a robot companion makes us feel good but leaves us somehow diminished?' (Turkle 2011). This takes us straight to the question of subjective and objective well-being, which I will shortly discuss in more detail.

In general, we use robots to solve problems or provide joy and entertainment. One reason for using robots is to alleviate loneliness, which was discussed in relation to the elderly. However, loneliness, it seems, is not exclusive to this group. Babies, children, adults and the old *all* seem to share a need for a certain degree of *companionship*. Erich Fromm (1994) writes about a need to relate to the outside world and that being deprived of this—living in loneliness and isolation—leads to *mental disintegration*.

Good thing, then, that in today's world, we are *never* really alone. If you look at the people on a bus, on a train, on the sidewalk, etc., you will see people getting on their phones as soon as the premonitions of boredom and loneliness strike. Connecting with the world, and their friends—never alone. The same function is filled by robot animals, robot pets, robot partners, etc. In our brave new world, is loneliness finally conquered?

There are, however, reasons to ask the question of how this affects us. One thing is that solitude may in fact be important for mental health, creativity and society in general. Mill (2004) writes of solitude as essential for depth of character. Storr (2005) describes how most new ideas come from solitude. Solitude and loneliness are not my prime concerns, but these are important issues related to how our current use of social technology affects us, and I refer to Turkle (2011) and Sætra (2018) for a more extensive account of these issues.

Then They Came for Me

When welfare technology is discussed in mainstream media, we usually get the sales pitch and a side note that ethics is important, too. We tend to get a reassurance that ethics is taken care of and that we should celebrate the innovative technologies that will lead to both more effective *and* better care. But social robots, and the ideas that let us deploy such robots in society, do not stop with the demented elderly. In this section I examine how the principles that have made us employ social robots in eldercare will allow us to do far more.

If we turn to the mainstream application of social AI, we might have a look at the *Tamagotchi,* which arrived in 1977. No body, but alive on a screen in an egg, these became a worldwide phenomenon that every kid, and many grown-ups, *had* to have. The Tamagotchis were shaped by the actions of their parents—they needed affection, care, feeding and attention. This they got – in spades! People were not simply entertained, as they formed relationships with these artificial beings. They *loved* them, were thrilled when they thrived, and even grieved when they died (Turkle 2011). Then there was the *Furby* and *My Real Baby*. So far, I have mostly discussed *toys*. With the *Aibo*, however, robots for adults arrived. Aibo is a *pet replacement*, with its '[r]ound, alluring eyes with a powerful pull, a cute, roly-poly form, moving around with infectious energy, and an identity …' (Sony 2018). Attachment to these artificial beings is not reserved for kids, as adults worry about the well-being of their electrified pets (Turkle 2011). If we were to believe the producers of these beings, it is not just humans that love the machines, but the machines love us too, as '[b]eing with people is what Aibo loves best' (Sony, 2018).

If you have a hard time putting your baby to sleep, Ewan the dream sheep comes to the rescue (SweetDreamers 2019). He senses when your baby is upset and provides soothing sounds and lights when your baby needs to calm down. Not very high intelligence, but intelligence embodied, nonetheless. SNOO has a different kind of body but also helps making babies sleep (HappiestBaby 2019). It is a robotic *cradle* that 'soothes upsets' by responding to upsets with both noise and motion, imitating the womb. What is the difference between parents applying such devices to ease the care for their babies and caretakers soothing the elderly with Paro?

As we see, they have come for the children, too. But have they come for us? Well, Roxxxy takes love to new heights, as she is a full-fledged companion robot (Hasse 2019). The idea of true robot companions for adults has garnered a lot of attention since Levy (2008) published his book on love and sex with robots more than 10 years ago (Cheok et al. 2016). Perhaps you liked the idea of Ewan and Snoo for sleep, but you cannot fit into a cradle and find it somewhat embarrassing to sleep with a sheep. Somnox is for you—the advanced Ewan equivalent for adults (Somnox 2019). It is a 'sleep robot' that lets you sleep *faster* and *longer*, making you more refreshed. Somnox might look like a pillow-seal hybrid, but it is designed for *affection*, as people are meant to hug it while sleeping, as it promotes a 'safe and secure feeling' (Somnox 2019). They *are* coming for us, too.

Ethics of Technology

There are many labels for the ethics of technology and machines, and AI ethics concerns the technologies of AI in general, including robots with AI. *Machine ethics* is concerned with how machines should act toward humans (Kochetkova 2014). Anderson and Anderson (2014), for example, write about the need for machines to act responsibly in accordance with ethical principles.

On the other hand, we have *robot ethics*, which focuses on the ethical behaviour of man in relation to machines. Making them, using them and treating them. One part of robot ethics is the question of whether or not robots have—or could or should have—*rights*. This is a question I do not discuss here, and I refer to Gunkel (2018) and Risse (2018) for a discussion of this topic. Another question I do not discuss is whether or not we can consider robots to be moral *agents*—capable of moral action in addition to simply being worthy of moral consideration by us (Gunkel 2012; Veruggio 2006).

I do, however, discuss the issue of *employing* robots, while I do not focus specifically on how we design them. While the ethics of technology is a vast and vibrant field, I have chosen to go back to some traditional concepts from the moral philosophy of old in my examination of the consequences of our application of technology, following Coeckelbergh's (2018) call for the use of political philosophy in the analysis of the implications of technology. The machines in themselves are not my main concern, as the questions I ask relate to how we *apply* them and what this application does to us.

These concepts are *hedonism, consequentialism* and *utilitarianism*. The reason I have chosen these terms is that I believe most justifications for the use of social robots are based on one, or all, of these ethical concepts.

Hedonism and Utilitarianism

Moyle et al. (2013) conducted a study showing how Paro the robot seal increased the *pleasure scores* for adults with dementia (compared with a reading group). If Paro gives pleasure, is this not a good reason to deploy him? The question at hand is really this: What sort of experiences are considered morally meaningful? The *subjective* experiences alone, or must we also consider the objective basis of our experiences? This is the question Turkle (2011) asked, when she asked if robots might in fact *diminish* us, even if they make us *feel good*.

Hedonism is a term used about the quest for *pleasure* and stems from the Greek word *hēdonē* (meaning *pleasure*). Hedonism comes in two forms: psychological and ethical. Psychological hedonism is a theory about human *motivation*, which says that pleasure is what motivates us. *Ethical* hedonism, on the other hand, is an ethical theory which places the achievement of pleasure, and avoidance of pain, at the centre of the evaluation of what is *good*. Jeremy Bentham (1996), which we shall return to shortly, stated that *pleasure* and *pain* are the governor of man, determining both what is *right* and *wrong* and what we endeavour to achieve by driving us toward pleasure and away from pain. This idea did not originate with Bentham, however, as Aristotle also saw happiness as the *final end* of human activity—'that for the sake of which every action is performed' (Cahn and Vitrano 2015).

Why, then, do I introduce hedonism in my treatment of social robots? Because the first proposition says that whatever promotes *well-being* is good and should be done. For the time being, let us assume that well-being and pleasure are closely related. If social robots promote happiness and relieve loneliness, pain and suffering, they are, from a hedonistic viewpoint, clearly good. What is interesting, however, is that many *other* things might also be justified by such an ethic, a point to which we shall return shortly.

I also consider the doctrine of utilitarianism, which is closely related to, but not necessarily identical with, hedonism. Utilitarianism is one of the most famous ethical doctrines in moral philosophy, and its foremost proponents are perhaps Mill and Bentham (1987). The aspect of most interest in this regard is the *consequentialist* part of the ethical theory of utilitarianism. Consequentialism is opposed to *deontological* ethics in that what determines what is *good* are the *consequences* that follow some action, not whether or not the action *in itself* can be considered right or wrong. For a consequentialist, the *ends* often justify the means, while this will not be the case for most deontologists, who evaluate the means by their own merits, regardless of consequences.

Utilitarianism is based on the *greatest happiness principle*, which says that the rightness of any action should be judged by the degree to which it promotes the greatest happiness for the greatest number of people (Mill and Bentham 1987). Consequentialism in the form of utilitarianism is introduced because it is closely related to the realist *second* proposition, which says that we should enact policies which promote the well-being of as many people as possible. This is, quite simply, a formulation of utilitarian ethic, and combined with the first proposition, we have a utilitarian ethic which embraces both hedonism and consequentialism.

The question I set out to answer is: Where do the principles that let us employ robot seals to the elderly lead us? In answering this, I rely on a somewhat simplified and generic understanding of the ethical principles involved. I recognize that various elaborations of the ethical principles I discuss might lead us in slightly different directions and I invite and encourage further debate on how different ethical doctrines will lead to different answers to the question at hand. My goal is to provide one such answer, but it is not the only one.

It might rightly be objected that I make a mistake when I equate happiness with well-being. I fully acknowledge that richer conceptions of well-being are possible and I will later return to a conception of the good that does not rely only on a person's subjective evaluations of their own condition. For now, however, well-being will be understood in the utilitarian sense of the balance of pleasure and pain a person experience.

Deception and the Willing Suspension of Disbelief

There are two facets of deception that are of interest in this setting. The first is that robots might fool humans into believing that they are real (Danaher 2020). This might be due to the sophistication of the robot or the inability to discriminate on the part of the human being. Babies, for example, are likely to be fooled easily, but so are a lot of the elderly with dementia. We might say that when this occurs, the robot has passed some individual's personal Turing test. Some people are far more vulnerable than others, but I argue that this is a difference in degree, not in kind.

The second facet of deception is that which occurs when we *know* that a machine is not real but cannot help ourselves from responding to it as if it was. As discussed earlier, robots are alive enough to trigger emotional responses in us, and these are often beyond both our perception and control. In this category I will also include what might be referred to by Coleridge's phrase 'the willing suspension of disbelief' (Jacobsen 1982).

Full Deception: Passing the Conscious Turing Test

The first category, here labelled *full deception*, occurs when a person believes an artificial entity to be something that it is not, namely, alive—usually in the form of another person or an animal.

One obvious example is the elderly with dementia that are unable to distinguish Paro the machine from a real animal. As Sharkey and Sharkey (2012a) note, their brains leave them vulnerable to deception. Passing the Turing tests of persons afflicted with old age and Alzheimer is not very difficult. That this actually happens is easily seen when reporters visist an eldercare facility, filming an old man heartily convinced that Paro is a real animal. He even comforts Paro when the reporter suggests that Paro is not real (Indreiten 2011; Flåm and Assev 2011).

This kind of deception is interesting because a deceived person may *subjectively* experience no harm nor foul. If the person will never realize that they were deceived, was any real harm done?

Another example would be very young children. The old and the young—the most vulnerable—are most easily fully deceived by technology. In the case of old people with dementia, there is little reason to expect them to suddenly become more aware of the difference between real and artificial and thus realize that they were deceived. Children, however, tend to grow up and *will* most likely realize what has happened at some point.

My approach to the possibility of full deception is theoretical, and it suffices to state that robots *today* are able to fully deceive some, and it seems likely that as we get more and more advanced robots, more and more people will be susceptible to full deception. I will not examine the possibility of being deceived by online chatbots, etc. in this article, but I *will* later consider a different kind of deception: that which occurs in Nozick's (2013) experience machine.

Partial Deception: Passing the Subconscious Turing Test

I call the second category *partial deception*. This includes all cases where a person, if asked, will respond that they know the machine not to be *real* in the sense that it is not a live, biotic being. The person in question does not, in their conscious thoughts, confuse the machine with a different kind of being. However, subconsciously, they will respond to it *as if* it was real. They are alive enough for us to respond to them *as if* they are real (Turkle 2011): bonding with it, responding emotionally to it, etc. It is of relatively little consequence if this process occurs *entirely* outside the consciousness of the person or if it is the result of a willing suspension of disbelief. What is *not* included here is the case where people treat these machines as kids treat dolls and project themselves upon them. For me to speak of deception, the machines must in some way make people treat them as *subjects*—something with a will of its own.

Sparrow (2002) discusses how the benefits people get from robot pets, social robots, etc., are predicated on believing, *at some level*, the animals to be real. If we are to get the benefits from, for example, animal-assisted therapy from a machine, this involves that we must be deceived, at least on an unconscious level.

The deception involved need not be *intentional* and will often result from anthropomorphism resulting from a machine imitating various aspects of life (Sharkey and Sharkey 2012a). While Sharkey and Sharkey (2011) suggest that people *enjoy* anthropomorphising technology, I mainly focus on how this happens even if we do not consciously do so. Turkle et al. (2006) relay the stories of how old people at times express an indifference to the reality (or lack thereof) of their new companions (Paro, in this instance), as they love them regardless.

The main question we must ask ourselves is: What do we think of deception? Those who take the stance that *no* deception is ever morally justified must, I argue, be opposed to all use of social robots (Sharkey and Sharkey 2012a). However, our propositions mention well-being and efficiency, so what happens when deception leads to both more well-being and more efficient care? There are trade-offs involved, and these are discussed in more detail in Sharkey and Sharkey (2010).

Nozick (2013) discusses *side constraints* as a way to prevent a utilitarian approach from justifying unreasonable sacrifices of individual's rights, and this is a concept I return to in the following discussion.

Dignity and Respect

A final ethical issue to consider is that of dignity. This is related to, but not identical with, the issue of deception.

Some view dementia as a second childhood and use this as an argument in favour of treating those suffering from it as children (Reisberg et al. 2002). Others object to this and will argue that the person, had she *not* been afflicted with age and dementia, might have objected to such treatment (Sharkey and Sharkey 2012a).

This idea is similar to that of Berlin's difference between an *empirical* and *authentic* self (Berlin 2002). Our empirical self is the characteristics that define us at any moment, which is the result of whatever conditions we have lived in up to this point. This means that the empirical self of a person being deprived of all exposure to arts, knowledge, language, etc. would be rather limited. The *authentic* self, on the other hand, is something akin to *potential* characteristics. This is what we might have been, or a self we might be able to achieve (Carter 1999).

One might argue that a person's *authentic* self survives the neural degradation that creates an infantile empirical self and that we might do well to assume that *respectful* treatment of any person entails an eye to what this person has been. If we consider only people's empirical selves, the road to employing technology in far more encompassing ways in order to maximize subjective well-being seems quite broad. And it is paved with good intentions.

Berlin suggests that a focus on some hypothetical authentic self might encourage authoritarian paternalism in an effort to make people better off despite of their *subjectively perceived* desires. The idea is that someone knows what is best for the person better than the person herself, making paternalistic intervention justifiable. Cayton (2006), however, believes that the *opposite*—the idea of a second childhood and a focus on the empirical self—opens the door to deceit and an authoritarian approach to care.

Carers and *relatives* have reported that they see doll therapy as 'demeaning, patronizing and inappropriate' (Mackenzie et al. 2006; James et al. 2006). If the patient is happy, however, I argue that it is the ethical principles, and not the opinion of some specific group of people not directly party to the affair, that must be used to stop such employment of technology. Sharkey and Wood (2014) consider the question of whether Paro is *demeaning* or *enabling*. Here, they argue that the benefits of Paro outweigh the negatives associated with deception and loss of dignity, arguing that we might see even more benefits, in which case they argue that employing Paro will become a moral imperative—a *should*.

This argument is important, as a decision to *not* employ a social robot in a situation where there is no real alternative *will* leave the potential recipient of robot care without any care. If we accept that social robots *do* have certain benefits, this will be a real cost that must be considered.

Where It Leads

> *Suppose there were an experience machine that would give you any experience you desired. Superduper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank, with electrodes attached to your brain. Should you plug into this machine for life, preprogramming your life's experiences? (*Nozick 2013*)*

If all that matters is reported happiness, Nozick's experience machine might just be the solution for all our woes (Nozick 2013). Experiences are programmed and tailor-made for each individual, and let us assume that once in the machine, we forget that anything else exists, and also forget that we ever chose to enter the machine. Why should we *not* use this machine?

The first proposition states that *a therapeutic device that increases a patient's mental and physical well-being should be employed*, while the second states that *when all else is equal, one should arrange society so as to improve the well-being of as many people as possible*.

If all that matters is subjective well-being, what objections could we have against such a life? The parallel to robotic companions should be clear, if we accept the premise that they make people better off because people falsely believe that they are alive. The old and demented might actually believe that the robotic seal is a live animal and this gives them joy and the feeling of companionship and connectedness.

Nozick himself mentions three reasons for refraining from using such a device, and we must consider if any of these might constitute a reason to stop using robots in care. The first is that we want to *actually* do certain things, and not just amass *experiences* of them. Nozick argues that our goal is not the experiences per se but going through the actions that produce such experiences (Nozick 2013). The experiences themselves are thus only of secondary interest. But, he asks, what would be the reason for desiring the activities themselves?

The second reason is that we want to 'be a certain way, to be a certain sort of person' (Nozick 2013). A person being deceived into thinking he is the king of England of old, amassing great experiences and showing great courage—*is* this really a person that has lived a full life, is he a courageous person, etc.? Nozick (2013) says that a person connected to a machine is next to *nothing*, and he compares choosing such a machine to *suicide*.

The third reason takes us back to the subjective vs. the objective, of authentic vs. artificial. The machine, Nozick (2013) argues, is limited to providing man-made experiences, which has no connection to anything *deeper*. It is merely a *simulation*, and not *real*. In this respect, the machine is like psychoactive drugs—providing experiences, but not *real* experiences. Such drugs are, some would suggest, merely 'local experience machines' (Nozick 2013).

How is the experience machine connected to the social robots, then? My argument is that such a machine is similar to employing social robots, which produces subjective well-being based on relationships and experiences produced by machines. If we suggest that deception is justified if subjective well-being is enhanced, how could we object to connecting the old with dementia to such machines?

We are here dealing with a hypothetical machine, and let us assume that (a) it provides life-like experiences, (b) it produces physiological stimulation in order to prevent physical decay and (c) it is programmed to provide the optimal and individually tailored environment conducive to produce the maximum amount of happiness. This means that it will provide the perfect mix of love, relationships, companionship, experiences, comfort, joy, thrills, etc. for every person connected.

If we agree that such a machine would provide the ideal way for old people to end their lives, what is stopping us from using it on babies? And young children? We have seen that we already use robots to put babies to sleep—doing our best to minimize their crying (increasing well-being) while enabling parents by way of letting them care for their children with minimum effort, or inconvenience.

What about young children? They already had Tamatgotchis, Furbys and the likes, so why not up the ante and provide some more advanced machines that will function as their personal coaches? They could prepare food for them, lecture them, entertain them, etc. Come to think of it, why not provide education through entertaining simulations in the Nozick machine? And while we are at it, and they are having the perfect time, what reason do we have for disconnecting them?

If you say, 'because parents need them', that is only because you believe that parents are not also connected to this machine. A machine that provides them with perfect relationships, and not the challenges, tears and frustrations of *real* children.

What is to stop us from connecting ourselves to such a machine? It could produce our ideal world. No loneliness, no suffering, ideal nutrition, ideal learning, no harm, and love without limits. Nozick suggests that we might simply want to 'live (an active verb) ourselves, in contact with reality' (Nozick 2013). But *why* would we prefer this over the experience-world?

Where does it all end? If we accept robots as authentic companions, and the relationships we form with them as *authentic*, we take a *posthumanist* stance. This involves the 'decentering' of man and has traditionally been associated with theories such as Næss' *ecosophy*, where all biotic beings are considered intrinsically valuable and part of our moral communities (Braidotti 2013; Næss 1989). If we look at the writings of Gunkel (2018) and others, we could imagine a broadening of our moral communities in ways which also include abiotic machines, such as robots. We could also take the stance that robots and technology *enhance* human beings, and, for example, if we view Nozick's experience machine as something that improves our human condition, we might see that as a transhumanist stance (Ferrando 2013).

Nozick (2013) does provide one possible way of escaping the slippery slope here mapped out, and that is his idea of *side constraints.* These are absolute individual rights—rights not to be killed, coerced, defrauded, and we might add *deceived* and treated with no respect—which are not amenable to the calculus of utilitarianism (Nagel 2013). These rights, or side constraints, can only be violated in order to prevent some person from violating the same rights in others, *not* to maximize some aggregated well-being, even if this was achievable (Nagel 2013). Nozick (2013) suggests that side constraints are not guides to action but *constraints* on action. This merely restates the Kantian categorical imperative, which implies that people should be treated as *ends*—not *means*. The rights of others—old, demented, healthy, young or insane—are not goals we should act on, but they do erect certain barriers to what we might justifiably do. If we have a desire to create a society in which *technology* promotes the *good society*, we must determine how such a society treats individuals (Griffy-Brown et al. 2018).

If we believe that people have individual rights demanding respectful treatment and non-deception, the arguments we currently use in favour of social robots will not suffice. All people have such rights, which imply that focusing on such rights enables us to critically consider *any* employment of social robots. The slippery slope argument is not the main reason to oppose the uncritical employment of social robots—it is merely a means to uncover and display the logic involved, which is problematic even if social robots were only used as a limited set of old people suffering from dementia.

Conclusion

Social robots in care have been shown to have positive effects on, for example, people with dementia. Despite these positive consequences, I have shown that there is a reason to have qualms about the uncritical employment of technology. This is, firstly, due to the effects social robots have on us and our ability to relate meaningfully with other people. If we use machines to escape all disagreeable situations, such as boredom, and to form perfect relationships where the *other* always caters to our own needs, we might *feel* good, but still end up diminished.

Secondly, there are the issues of deception and dignity. Let us say we felt happy being treated liked children and pandered and cared for in ways that we would abhor had we observed ourselves from a position of full mental capacity. Should we as a society then say that this *subjective* well-being justifies deceit and ignoble treatment of people?

I argue that the principles of hedonism and utilitarianism can easily lead to a defence of the use of social robots. What is even more interesting, however, is that these very principles might take us to a future far more radical than one in which our elderly get pacifiers in the form of robot seals.

If subjective well-being is all that matters, and consequences—not absolute individual rights—are our yardstick for right or wrong, technology can be employed to pacify and satisfy us all. There is little to suggest that if we use robots to create joy in our elders' 'second childhood', why not do the same in the first? And why not give robot companions to our young—to tutor them, care for them and entertain them—creating massive benefits for society and joy for the children? And grownups? We no longer need to suffer the intricacies of other people. Robot companions can provide what we need also in terms of intimate relationships. Again, great for society, as such relationships are important for mental health. Not to mention the violence, suffering and sickness caused by broken relationships. Come to think of it, real relationships are so dangerous that they should most likely be banned in order to protect adults, the children and our societies.

But what is left, if we accept the employment of technology in such ways? I argue that there is little to stop us from advocating in favour of connecting us all to something akin to Nozick's experience machine, if hedonism and utilitarianism are all that matters. Happiness all around—perfect nutrition and health through scientific mental and physical stimulation. First, they came for the old and demented. Then, they came for the rest of us.

We do, however, have the opportunity to say 'stop'. We might temper consequentialism with Nozick's side constraints, demanding that people's right to be treated with a certain respect be enforced. First, they came for the old and demented, and we said *stop*.

Funding Information Open Access funding provided by Ostfold University College.

References

Anderson, L. S. & Anderson, M. (2014). Towards a principle-based healthcare agent. In S.P. van Rysewyk and M. Pontier (Eds.), *Machine medical ethics*. Springer.

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters, 2*(3), 412–414.

Bemelmans, R., Gelderblom, G. J., Jonker, P., & De Witte, L. (2012). Socially assistive robots in elderly care: a systematic review into effects and effectiveness. *Journal of the American Medical Directors Association, 13*(2), 114–120.

Bendel, O. (2014). Surgical, therapeutic, nursing and sex robots in machine and information ethics. In S.P. van Rysewyk and M. Pontier (Eds.), *Machine medical ethics*. Springer.

Bentham, J. (1996). *The collected works of Jeremy Bentham: An introduction to the principles of morals and legislation*. Clarendon Press.

Berlin, I. (2002). Two concepts of liberty. In H. Hardy (Ed.), *Liberty*. Oxford: Oxford University Press.

Bernabei, V., De Ronchi, D., La Ferla, T., Moretti, F., Tonelli, L., Ferrari, B., et al. (2013). Animal-assisted interventions for elderly patients affected by dementia or psychiatric disorders: a review. *Journal of Psychiatric Research, 47*, 762–773. https://doi.org/10.1016/j.jpsychires.2012.12.014.

Bodenhagen, L., Suvei, S. D., Juel, W. K., Brander, E., & Krüger, N. (2019). Robot technology for future welfare: meeting upcoming societal challenges–an outlook with offset in the development in Scandinavia. *Health and Technology, 9*(3), 197–218.

Braidotti, R. (2013). *The Posthuman*. Cambridge: Polity.

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B. & Anderson, H. (2018). The malicious use of artificial intelligence: forecasting, prevention, and mitigation. arXiv preprint arXiv:1802.07228.

Burr, C., Taddeo, M., & Floridi, L. (2019). The ethics of digital well-being: A thematic review. Available at SSRN 3338441.

Cahn, S. M., & Vitrano, C. (2015). *Happiness and goodness: philosophical reflections on living well*. Columbia University Press.

Carter, I. (1999). *A measure of freedom*. Oxford University Press.

Cayton, H. (2006). 17 from childhood to childhood? Autonomy and dependence through the ages of life. Dementia Mind, Meaning, and the Person, 277.

Cheok, A. D., Levy, D., & Karunanayaka, K. (2016). Lovotics: love and sex with robots. In *Emotion in Games* (pp. 303–328). Cham: Springer.

Chiberska, D. (2018). The use of robotic animals in dementia care: challenges and ethical dilemmas. Mental Health Practice, 22(3).

Coeckelbergh, M. (2018). Technology and the good society: a polemical essay on social ontology, political principles, and responsibility for technology. *Technology in Society, 52*, 4–9.

Cominelli, L., Mazzei, D., & De Rossi, D. E. (2018). SEAI: Social emotional artificial intelligence based on Damasio's theory of mind. *Frontiers in Robotics and AI, 5*, 6.

Damasio, A. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Quill.

Damasio, A. (2003). *Looking for Spinoza: joy, sorrow, and the feeling brain*. Orlando: Harcourt.

Damasio, A. (2018). *The strange order of things*. New York: Pantheon Books.

Danaher, J. (2020). Robot betrayal: a guide to the ethics of robotic deception. *Ethics and Information Technology*. https://doi.org/10.1007/s10676-019-09520-3.

Ding, Y., Sohn, J. H., Kawczynski, M. G., Trivedi, H., Harnish, R., Jenkins, N. W., Lituiev, D., Copeland, T. P., Aboian, M. S., Mari Aparici, C., & Behr, S. C. (2018). A deep learning model to predict a diagnosis of Alzheimer disease by using 18F-FDG PET of the brain. *Radiology, 290*(2), 456–464.

Döring, N., & Poeschl, S. (2019). Love and sex with robots: a content analysis of media representations. *International Journal of Social Robotics*, 1–13.

Downes, M. J., Dean, R., & Bath-Hextall, F. J. (2013). *Animal-assisted therapy for people with serious mental illness (protocol), Cochrane database of systematic reviews*. Loughborough: John Wiley & Sons.

Ferrando, F. (2013). Posthumanism, transhumanism, antihumanism, metahumanism, and new materialisms. *Existenz, 8*(2), 26–32.

Flåm, K. & M. Assev. (2011). Ekspertforslag skaper reaksjoner: Vil gi robotomsorg til eldre. VG. Retrieved from https://www.vg.no/nyheter/innenriks/i/pzwMG/ekspertforslag-skaper-reaksjoner-vil-gi-robotomsorg-til-eldre

Fromm, E. (1994). *Escape from freedom*. New York: Henry Holt and Company.

Griffy-Brown, C., Earp, B. D., & Rosas, O. (2018). Technology and the good society. *Technology in Society, 52*, 1–3.

Gunkel, D. J. (2012). *The machine question: critical perspectives on AI, robots, and ethics*. MIT Press.

Gunkel, D. J. (2018). *Robot rights*. MIT Press.

Happiest Baby. (2019). Snoo: smart sleeper. Retrieved from https://www.happiestbaby.com

Hasse, C. (2019). The Vitruvian robot. *AI & SOCIETY, 34*(1), 91–93.

Hobbes, T. (1946). *Leviathan*. London: Basil Blackwell.

Indreiten, A. B. (2011). Robotselen Paro kjenner igjen Sverre. NRK. Retrieved from https://www.nrk.no/vestfold/robotomsorg-for-demente-1.7769505

Jacobsen, M. (1982). Looking for literary space: the willing suspension of disbelief re-visited. *Research in the Teaching of English*, 21–38.

James, I. A., Mackenzie, L., & Mukaetova-Ladinska, E. (2006). Doll use in care homes for people with dementia. *International Journal of Geriatric Psychiatry: A Journal of the Psychiatry of Late Life and Allied Sciences, 21*(11), 1093–1098.

Johnston, O., & Thomas, F. (1995). *The illusion of life: Disney Animation* (pp. 306–312). New York: Hyperion.

Keynes, J. M. (1923). *A tract on monetary reform*. London: Macmillan.

Kochetkova, T. (2014). An overview of machine medical ethics. In S.P. van Rysewyk and M. Pontier (Eds.), *Machine medical ethics*. Springer.

Latikka, R., Turja, T., & Oksanen, A. (2019). Self-efficacy and acceptance of robots. *Computers in Human Behavior, 93*, 157–163.

Levy, D. (2008). *Love and sex with robots: the evolution of human-sex relationships*. New York: Harper Perennial.

Lockett, J. (2017). World's first brothel staffed entirely by robot sex workers now looking for investors to go global. The Sun. Retrieved from https://www.thesun.co.uk/news/4131258/worlds-first-brothel-staffed-entirely-by-robot-sex-workers-now-looking-for-investors-to-go-global/

Machiavelli, N. (2003). *The Prince*. New York: Bantam Books.

Mackenzie, L., James, I. A., Morse, R., Mukaetova-Ladinska, E., & Reichelt, F. K. (2006). A pilot study on the use of dolls for people with dementia. *Age and Ageing, 35*(4), 441–444.

Metzler, T. A., Lewis, L. M., & Pope, L. C. (2016). Could robots become authentic companions in nursing care? *Nursing Philosophy, 17*(1), 36–48.

Mill, J. S. (2004). *Principles of political economy*. New York: Prometheus Books.

Mill, J. S., & Bentham, J. (1987). *Utilitarianism and other essays*. Penguin UK.

Mordoch, E., Osterreicher, A., Guse, L., Roger, K., & Thompson, G. (2013). Use of social commitment robots in the care of elderly people with dementia: a literature review. *Maturitas, 74*(1), 14–20.

Moyle, W., Cooke, M., Beattie, E., Jones, C., Klein, B., Cook, G., & Gray, C. (2013). Exploring the effect of companion robots on emotional expression in older adults with dementia: a pilot randomized controlled trial. *Journal of Gerontological Nursing*.

Næss, A. (1989). *Ecology, community and lifestyle*. Cambridge: Cambridge University Press.

Nagel, T. (2013). Foreword by Thomas Nagel. In Nozick, R. (2013). *Anarchy, state, and utopia*. New York: Basic Books.

NOU 2011:11. (2011). *Innovation in the Care Services*. Retrieved from https://www.regjeringen.no/en/dokumenter/nou-2011-11/id646812/

Nozick, R. (2013). *Anarchy, state, and utopia*. New York: Basic Books.

Paro Robots. (2018). *Paro therapeutic robot*. Retrieved from http://www.parorobots.com

Poulsen, A., &Burmeister, O. K. (2019). Overcoming carer shortages with care robots: dynamic value trade-offs 805 in run-time. Australasian Journal of Information Systems, 23.

Prince, M., Guerchet, M., & Prina, M. (2013). *Policy brief for heads of government: the global impact of dementia 2013–2050*. London: Alzheimer's disease international (ADI).

RealBotix. (2019). realdoll^{x}. Retrieved from https://www.realdollx.ai

Reisberg, B., Franssen, E. H., Souren, L. E., Auer, S. R., Akram, I., & Kenowsky, S. (2002). Evidence and mechanisms of retrogenesis in Alzheimer's and other dementias: management and treatment import. *American Journal of Alzheimer's Disease & Other Dementias®, 17*(4), 202–212.

Risse, M. (2018). *Human rights and artificial intelligence: an urgently needed agenda*. Cambridge: Carr Centre for Human Rights Policy.

Roger, K., Guse, L., Mordoch, E., & Osterreicher, A. (2012). Social commitment robots and dementia. *Canadian Journal on Aging/La Revue canadienne du vieillissement, 31*(1), 87–94.

Sætra, H. S. (2018). The ghost in the machine: being human in the age of AI and machine learning. *Human Arenas., 2*(1), 60–78. https://doi.org/10.1007/s42087-018-0039-1.

Sætra, H. S. (2020). The parasitic nature of social AI: Sharing Minds with the Mindless. *Integrative Psychological and Behavioral Science*. https://doi.org/10.1007/s12124-020-09523-6.

Savela, N., Turja, T., & Oksanen, A. (2018). Social acceptance of robots in different occupational fields: a systematic review. *International Journal of Social Robotics, 10*(4), 493–502.

Scheutz, M., & Arnold, T. (2016). Are we ready for sex robots? In *the Eleventh ACM/IEEE International Conference on Human Robot Interaction* (pp. 351–358). IEEE Press.

Schwab, K. (2019). How googly eyes solved one of today's trickiest UX problems. Fast Company. Retrieved from https://www.fastcompany.com/90395110/how-googly-eyes-solved-one-of-todays-trickiest-ux-problems

Sharkey, A., & Sharkey, N. (2010). Living with robots: ethical tradeoffs in eldercare. In Wilks, Y. (Ed.). (2010). *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues* (Vol. 8). John Benjamins Publishing.

Sharkey, A., & Sharkey, N. (2011). Children, the elderly, and interactive robots. *IEEE Robotics & Automation Magazine, 18*(1), 32–38.

Sharkey, N., & Sharkey, A. (2012a). The eldercare factory. *Gerontology, 58*(3), 282–288.

Sharkey, A., & Sharkey, N. (2012b). Granny and the robots: ethical issues in robot care for the elderly. *Ethics and Information Technology, 14*(1), 27–40.

Sharkey, A., & Wood, N. (2014). The Paro seal robot: demeaning or enabling. In *Proceedings of AISB* (Vol. 36).

Soler, V., Meritxell, L. A.-O., Rodríguez, J. O., Rebolledo, C. M., Muñoz, A. P., Pérez, I. R., Ruiz, E. O., et al. (2015). Social robots in advanced dementia. *Frontiers in Aging Neuroscience, 7*, 133.

Sony. (2018). aibo. Retrieved from https://aibo.sony.jp/en/

Somnox. (2019). Meet somnox. Retrieved from https://meetsomnox.com

Sparrow, R. (2002). The march of the robot dogs. *Ethics and Information Technology, 4*(4), 305–318.

Storr, A. (2005). *Solitude: a return to the self.* New York: Free Press.

SweetDreamers. (2019). Ewan the dream sheep. Retrieved from https://sweetdreamers.co.uk/product-category/ewan-the-dream-sheep-baby-sleep-aid-and-soother/

Turing, A. M. (2009). Computing machinery and intelligence. In R. Epstein, G. Roberts, & G. Beber (Eds.), *Parsing the Turing Test* (pp. 23–65). Dordrecht: Springer.

Turkle, S. (2011). *Alone together: why we expect more from technology and less from each other*. New York: Basic Books.

Turkle, S., Taggart, W., Kidd, C. D., & Dasté, O. (2006). Relational artifacts with children and elders: the complexities of cybercompanionship. *Connection Science, 18*(4), 347–361.

Vandemeulebroucke, T., Dierckx de Casterlé, B., Welbergen, L., Massart, M., & Gastmans, C. (2019). The ethics of socially assistive robots in aged care. A focus group study with older adults in Flanders, Belgium. *The Journals of Gerontology: Series B*.

Veruggio, G. (2006). The euron roboethics roadmap. In *Humanoid Robots, 2006 6th IEEE-RAS International Conference on Humanoid Robots* (pp. 612–617). IEEE.

Wada, K., Shibata, T., Musha, T., & Kimura, S. (2008). Robot therapy for elders affected by dementia. *IEEE Engineering in Medicine and Biology Magazine, 27*(4), 53–60.

Winfield, A. (2012). *Robotics: a very short introduction*. Oxford: Oxford University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.