Technology in Society 63 (2020) 101383

Contents lists available at ScienceDirect

Technology in Society

journal homepage: http://www.elsevier.com/locate/techsoc

The foundations of a policy for the use of social robots in care

Henrik Skaug Sætra

Østfold University College, Remmen, 1757, Halden, Norway

ARTICLE INFO

Keywords:
Care
Social robots
Ethics
Utilitarianism
Policy

ABSTRACT

Should we deploy social robots in care settings? This question, asked from a policy standpoint, requires that we understand the potential benefits and downsides of deploying social robots in care situations. Potential benefits could include increased efficiency, increased welfare, physiological and psychological benefits, and experienced satisfaction. There are, however, important objections to the use of social robots in care. These include the possibility that relations with robots can potentially displace human contact, that these relations could be harmful, that robot care is undignified and disrespectful, and that social robots are deceptive. I propose a framework for evaluating all these arguments in terms of three aspects of care: *structure, process,* and *outcome*. I then highlight the main ethical considerations that have to be made in order to untangle the web of pros and cons of social robots in care as these pros and cons are related the trade-offs regarding quantity and quality of care, process and outcome, and objective and subjective outcomes.

1. Introduction

Should we deploy social robots in care settings? In this article I search for the policy desiderata for social robots in care. This endeavour is clearly related to, but also distinct from, the question of how social robots in care should behave. It is also related to fundamental questions of human-robot interaction, and such questions as what happens when we relate socially to robots. By drawing on the literature on social robots, political theory, and ethics, I provide the grounds for policy formation on the use of social robots in care.

On the one hand, there are many potential benefits to be had from a social robot. Economists and politicians see important demographic changes looming, which require a new supply of carers for the increasing population of elderly [1]. Robotic engineers see the benefits of providing assistance, entertainment, and companionship for people by equipping robots with social artificial intelligence, and researchers in the field of care have also found evidence of benefits, both physiological and psychological, of using social robots in care situations [2].

On the other hand, there are potential negative consequences of deploying social robots in care. First of all, they have the potential to displace human beings, and human beings may in fact be an important factor in providing good care. Social robots that relate to human beings may also be harmful if the relations they form with human beings leave the humans worse off. Finally, it can be argued that any benefit that is observed in the use of social robots is achieved through deception and undignified treatment of the cared-for.

In order to evaluate these pros and cons, we need a better understanding of how social robots impact the *quality* and *quantity* of care. I use and modify a framework for assessing the quality of care though the examination of the *structure, process, and outcome* of care [3]. I show how social robots have both positive *and* negative consequences for all three aspects of the quality of care and how a basic ethical framework can be used to untangle the various effects of social robots on the quality of care.

I show that the answer to the first question in this introduction depends on whether or not we emphasise the *outcome* or *process* of care, and whether or not we decide to focus on *individual* recipients of care or the *general* level of care in society. We also have to consider whether those being cared for are the best judges of what a good outcome of care entails.

In sum, I show that the first question is deeply political and can best be answered by the tools of political philosophy and ethics. I share Coeckelbergh's [4] concern about *implicit assumptions* at play in the analysis of technology as well as his suggestions that political theory has much to offer in debates about foundational political principles. The answer to the question of whether or not, or how to, use social robots in care situations is dependent on the definition of a good *society* and the role of technology in such a society [5]. This is an attempt to highlight the main ethical and political issues involved for a broader audience, in particular those involved in politics and the formation of healthcare policy.

E-mail address: Henrik.satra@hiof.no.

https://doi.org/10.1016/j.techsoc.2020.101383
Received 7 April 2020; Received in revised form 24 August 2020; Accepted 4 September 2020
Available online 6 September 2020

2. Social robots in care

Robots in care has garnered attention and some commercial success over time, partly because of certain demographic developments observed in modern societies [6,7]. For example, people grow older, and many of the elderly face challenges such as dementia and a general need for *care* [1]. Upholding the ratio of carers to elderly is perceived as a great challenge, and it is hoped that robots can meet some of the needs created by a shortage of carers [8,9]. According to Fong et al. [10], *social* robots will, for example, 'assist in health care, rehabilitation, and therapy'.

The social robots I discuss in this article are created to interact with people and fulfil certain functions that people normally seek in human or animal social companions. Robots, it is argued, may require *social* functionality in order to operate properly when integrated in human social settings – particularly when human-robot interaction and cooperation is the goal [11]. When the term *social robot* is used in this article, it refers to robots that are socially *receptive* but not *sociable* robots, as the latter are considered to have 'their own internal goals and motivations' according to Breazeal [12].

While social robots have a physical presence and *can* be endowed with the ability to provide physical assistance such as lifting, holding, feeding, etc, I will only consider the *social* aspect of such robots. Socially *assistive* robots are defined by Feil-Seifer and Mataric [13] as the intersection between socially interactive robots and assistive robots. The aspects of robots here analysed can include *cognitive,* but not *physical,* assistance, which means that I do not discuss socially assistive robots in their entirety. The type of social robot in question cannot manipulate objects or provide any form of physical assistance. It is not a lifter, washer, feeder, mover or butler. Neither will it be equipped with sensors and other capabilities primarily intended to assist patients with physical monitoring and assistance [14]. While the robot cannot assist physically, it can provide emotional and cognitive assistance [15]. These three forms of assistance are what Johansson-Pajala et al. [16] consider when they speak of *care robots*, and the *physical* aspect is excluded from the current analysis. According to their definition of a *personal care robot*, which specifies 'robotics that improve the quality of life of humans, on a non-medical basis', the social robots included in the analysis could still be considered such robots [16]. Examples of robots that together fulfil the same functions as the social robots here analysed are Buddy, Paro, Leka, and Cutii [17–20].

I consider the type of robots discussed by Sharkey and Sharkey [21] and Bradwell et al. [7], which provides *companionship* in addition to basic cognitive assistance. Included in the analysis are current or near-future social robots, and their intelligence is as a consequence limited. They can interact rather convincingly with other beings, primarily through spoken and body language. There is, however, no *strong AI* or *general* intelligence present, and these robots are not assumed to *understand* anything in the human sense of the word [22].

Coeckelbergh [23] and Sparrow [24] have both described thought experiments concerning robot eldercare. They describe counter-factual scenarios where the elderly are cared for in facilities where robots perform all necessary duties. This article is an attempt to identify the core ethical issues invoked by the use of *social* robots in care, and to develop the policy desiderata that follow from the consideration of these issues.

The question, for those developing policies regarding the use of social robots in care, is whether or not social robots should be deployed. An examination of the pros and cons is in order — which arguments can be made for and against the deployment of social robots?

3. Why deploy social robots?

A premise for considering the deployment of social robots is that it must be based on the provision of a tangible benefit. There are two main categories of such benefits, namely the increase in the *quality* or *efficiency* of care [25]. I will consider both categories, each of which contains several different arguments. I must also note that in the following I examine findings from multiple types of robots (animal-like and humanoid) as well as various types of patient groups (elderly, with or without dementia, the mentally ill, etc.). This is not to suggest that findings are fully transferable between these categories, and I simply use this examination to build a strong case for *potential* benefits of deploying social robots in care situations.

3.1. Increased quality of care

Quality of care is an issue often discussed but not quite as often defined. For example, Johansson-Pajala et al. [16] speak of the need to 'facilitate the acceptance and normalisation of care robots in elder care for exploiting the benefits such as increased quality in elder care and improved independence in everyday life'. The fact that robots can improve quality in eldercare is taken as a given, but we must delve a bit deeper into both the concept of *care* and *quality*.

Care involves situations in which someone is dependent on assistance from others for various reasons, particularly in old age. The provision of care is often assumed to be based on a *moral claim*, either involving individuals close to a person or society at large [26].

Many definitions of *care* have been attempted, some very broad and some very specific [26]. One definition by Bubeck [27] states that '[c]aring for is the meeting of the needs of one person by another person, where face-to-face interaction between carer and cared-for is a crucial element of the overall activity and where the need is of such a nature that it cannot possibly be met by the person in need herself'. Care, then, is based on the presence of certain needs— 'those which make us dependent on others' [27]. Bowden [28] further states that caring 'expresses ethically significant ways in which we matter to each other, transforming interpersonal relatedness into something beyond ontological necessity or brute survival'.

Such definitions illuminate the problem of the very concept of a *care robot*. The latter definition would, I argue, completely rule out the possibility of care robots, but the first may be amended to allow for the provision of robot care. If we simply state that care involves the meeting of the needs of a person but skip the demand that the carer be a *person*, robots can care. Furthermore, Bubeck [27] states that face-to-face interaction between the carer and the cared-for is crucial, and I will *assume* that robots are able to interact face-to-face, while recognising that this is a controversial assumption.[1] Meacham and Studley [31] similarly ask if a robot can *care*. They believe that it can, as they see care as based on the *external* rather than the *internal* aspects of the carer and the environment.

In developing policy desiderata for the use of social robots, I begin with Donabedian [3] and his three main categories for evaluating quality of care: *structure*, *process*, and *outcome,* as shown in Fig. 1.

Structure involves human and physical resources, such as facilities and staff. *Process* involves an examination of *how* care is both attained and provided. *Outcome* encompasses measures of the *health status* after care is provided, which involves both objective measures of health *and* patients' knowledge, behaviour, and satisfaction. This framework lets us evaluate the achievement of good outcomes through bad processes, good process that do not have the desired outcome, and both *objective* and *subjective* evaluations of the achieved outcomes. We will now turn to some of the potential benefits of social robots in care, as discussed in the literature on social robots.

3.1.1. Improved physiological factors

In a systematic study of the effects of effects of 'socially assistive robots', Bemelmans, et al. [25] found positive effects on physiological

[1] See Gunkel [29] for an elaboration of Levinas's emphasis on *face-to-face* relations related to robots [30].

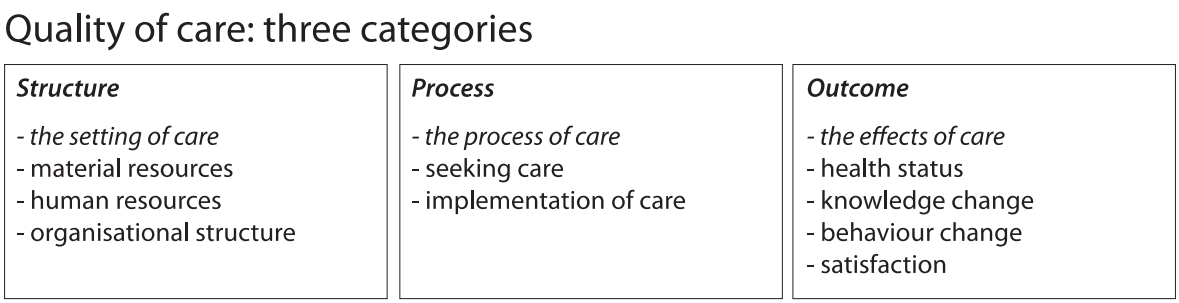

Fig. 1. Aspects of quality of care [3].

factors, such as stress levels, in the cared-for. This is presumably achieved through the same kind of mechanism as those involved in animal-assisted therapy (AAT). Such therapy has been shown to calm patients and have various other positive effects [32].

Furthermore, the cognitive assistance provided by a social robot could, for example, lead to a healthier lifestyle by way of assisting in healthy eating, suggesting activities, managing medication, etc. Such factors could lead to objective improvements in care outcome. Patients could become healthier, which could be considered an objective improvement in care outcome.

3.1.2. Improved psychological factors

A direct consequence of improved physiological health could be improved *psychological* health. This would be an indirect benefit of social robots. There are, however, indications that social robots could also improve psychological health *directly*.

Bemelmans et al. [25] found positive effects from socially assistive robots on several psychological factors, including *mood, loneliness, social connections*, and *communication*. These benefits are also associated with AAT, and one of the selling points of social robots, such as Paro the robotic seal, is that they *do* provide the therapeutic effects of animal-assisted therapy [18]. The heralded benefits of such robots are great, and includes stress reduction, stimulation of interaction between carers and the cared-for, improved relaxation and motivation, and improved socialisation [18].

Furthermore, Moyle, et al. [33] show that companion robots increased the *pleasure scores* of the cared-for. This means that a potential benefit of social robots is that they make the cared-for *happier* in the sense that they report higher experienced pleasure. If the care we provide causes people to report high satisfaction, that is one of the subjective *outcome* factors involved in evaluating the quality of care.

I will also note that we need not assume that human (or animal) therapy is the measure to which care by robots must aspire. In an examination of potential benefits of social robots, we must accept the possibility that robots with advanced artificial intelligence may provide *better* care. This could be the case if the cared-for preferred not to expose their vulnerabilities to other human beings. Some might, for example, prefer to have a robot monitor them, for example in the shower and in the bathroom, as opposed to having a human caretaker [34]. Another possibility is that robots, armed with advanced artificial intelligence, could provide more advanced and effective care than human caretakers could.

3.1.3. Empowerment

While robot help with physical tasks is not considered, robots can provide cognitive assistance. This means that they can provide guidance regarding the proper time for and amount of medication, they can assist in nudging the cared-for towards a somewhat more active lifestyle, they can remind the cared-for who their family members are, where they left important items, etc.

This kind of assistance means that a social robot can *empower*, and 'facilitate and enhance older people's autonomy' [24]. It would let the cared-for be more independent (of other *people*) than they would otherwise be, and it could also allow the cared-for to 'age-in-place', which could be beneficial both for the cared-for and those who pay for the alternative—institutionalised care [35]. In addition, they can perform safety functions such as making sure that the oven is turned off, and even monitor and report if the cared-for should fall, have a heart attack, etc. The latter functions, however, are not part of the social aspect of social robots.

3.2. Improved efficiency of care

Care is costly, no matter how we look at it. On a social level, it requires vast amounts of resources, both human and material. It is also necessary, but most tend to agree that more efficient care, *ceteris paribus*, is preferable to the opposite. If not for cynical reasons and the quest for profit, increased efficiency would allow us to provide more care.

3.2.1. Monetary efficiency

One claimed benefit of care robots is that they could provide care at a lower cost. Sparrow and Sparrow [36] are sceptical of such claims, and point out that robots with the required sophistication to perform the duties promised are quite expensive. However, let us *assume* that social robots without physical assistance capabilities can be produced relatively cheaply now or in the near future. If we were to achieve a slightly reduced demand for human care by deploying such robots, we will assume that this leads to monetary efficiency.

All sorts of things may be wrong with the all-robotic facility imagined by Sparrow [24], but we cannot automatically assume that the care provided in such facilities would be worse than the alternative. One of the main reasons for pursuing economic efficiency is that we will be able to provide *more* care for *less* money. This means that poor societies could provide care for those that would otherwise not be cared for, and even in rich societies care could be provided for more people for longer periods of time and in new situations [37].

We cannot assume that the sole goal of making care more efficient is either to increase profits from care or use the money saved from care in other sectors of society. Increased monetary efficiency is thus considered to be a benefit.

3.2.2. Increased flexibility and human-robot cooperation

Another potential benefit of social robots is that we might *change* the way care is provided. Social robots need not *replace* human caretakers, but they could assist them [38]. While social robots provide certain tasks, human caretakers could provide other tasks for which they previously did not have time. Care could also be provided *faster* if humans and machines cooperated on the performance of tasks.

3.2.3. Robot-specific opportunities for care

Another benefit is seen in situations in which robots have clear and obvious advantages over human beings. One such settings would be in hazardous situations, for example in poisonous or radioactive settings. In a hypothetical setting where care would either have to be provided in

such a setting or not at all, a social robot would obviously provide benefits. This would be akin to the robots that disarm bombs—taking advantage of their immunity, durability, and expendability in order to perform tasks that humans either cannot or will not perform.

In the COVID-19 situation in 2020, the elderly were particularly at risk and were isolated in care facilities in large parts of the world. Only their caretakers could enter the facilities, and while necessary, this also involves the risk of infected caretakers bringing the virus into the facilities. A social robot alone would not fix this situation, as it could not provide any physical assistance. However, it *could* provide companionship at a time of particular need, in a situation where the cared-for were not allowed to meet their friends and relatives.

3.3. The benefits summarised

We have seen that there are several potential benefits to be had from social robots in care. As to the *structural* aspects of quality of care, increased efficiency, and particularly new opportunities derived from such robots' unique nature could lead to increased structural quality of care. This is because structural conditions are improved by freeing up resources, opening up for new use of existing resources, and creating new arenas of care.

The main advantages, however, are related to the *outcome* aspect of care. Both objective and psychological factors could be improved by deploying social robots. Objective and subjective health status could be improved. These could, but need not, be associated.

Lastly, we could also see improvements in the *process* aspect of care. One possibility is that robots replace human caretakers in situations where the cared-for feel particularly vulnerable, but also through more *efficient* care in terms of the access to care and the speed of the provision of specific tasks. All aspects of quality of care could potentially be positively impacted by the deployment of social robots.

4. Objections to social robots in care

Before we decide to deploy social robots, we must also examine the possible negative consequences of using such robots in care settings. As with the benefits, I here consider potential negative consequences drawn from the literature on social robot and robots in general, and I do not limit the discussion to the use of robots in *care*. This approach provides us with a broad arsenal of possible objections considered relevant for the use of social robots in care settings.

Two main concerns related to human-robot interaction are *deception* and *authenticity* [38]. In addition to these, there are questions of *dignity*, *respect*, and *recognition* [21,24]. Stahl and Coeckelbergh [39] consider the effects of healthcare robotics in general, which includes robots providing physical assistance. While they consider factors such as de-humanisation and deception, they focus more generally on issues related to the use of information technology and robotics, such as job loss and data ethics (including issues of privacy and data protection). My task is aimed at a specific subset of healthcare robotics; the more general issues related to information technology and robots are not considered here.

I therefore do not consider issues of *machine ethics*, understood as the examination of how robots should act *as subjects* [38,40]. Machine ethics is of great importance for guiding and evaluating the development of social robots, and I support the call for more ethical involvement in innovation processes [39]. In this article, however, I search for the policy desiderata for social robot in care, and this endeavour is clearly related to, but also distinct from, the question of how social robots should behave.[2]

[2] Kochetkova [9] provides an account of the main issues of *machine medical ethics:* "computability, robots as autonomous moral agents, and the relation between top-down, bottom-up and hybrid theoretical approaches".

4.1. Authenticity and human relationships

A robot is not a biotic being, and there is some concern about relations with social robots not being *authentic* or *meaningful* [36,41]. The objection is based on two related, but different, concerns. The first is that relationships with robots may *change* human beings. The second is that human beings have a *need* for human contact and relationships, and that the introduction of social robots will crowd out human contact.

While the relationships between humans and robots have been studied for a long time, there is still a lack of certainty with regard to the consequences of forming relations with robots [21,29]. Turkle [41] examines how easily we form social bonds with machines and how much we seem to enjoy them, but asks, 'What if a robot companion makes us feel good but leaves us somehow diminished?' This strikes at the core of the benefit of differentiating between *objective* and *subjective* outcomes.

This objection is also related to a purported need for human contact. One way of approaching this problem is to consider Maslow's hierarchy of needs and what he calls *social* or *belonging needs* [42]. Some also speak of a need to relate to others and that a lack of such relations leads to mental disintegration [43]. Such objections assume that robots *cannot* fulfil core companion skills, but there is still an ongoing debate about what *authenticity* really refers to and not least to the potential of social AI to develop into what we might perceive as true social companions [44,45].

4.2. Dignity and respect

A related factor is what Sparrow [24] calls the human need for recognition and respect, 'which robots cannot provide'. Recognition 'consists in the enjoyment of social relations that acknowledge us in our particularity and as valued members of a community', and respect 'consists in social and political relationships wherein our ends are granted equal weight to those of others in the community' [24]. The question of equal weight to each person's ends and goals is further discussed in section 6, and the concept of *recognition* is clearly also related to the question of authenticity.

Another perspective on the value of *respect* is the concept of *dignity*. Sharkey and Sharkey [21] write about the prospect of an erosion of the dignity of the cared-for, and relate this to the idea that it is undignified for people to be cared for and kept happy through a process that a person 'might have abhorred' had they had full or former mental capacity.

Some see those with dementia as passing through a *second childhood* and thus see this as reason to treat them like children [46]. Once more, we see that there is a fundamental concern about the difference between *objective* and *subjective* evaluation of outcomes. A person with dementia might be perfectly happy playing with a robotic seal—even believing that it is real—but some could argue that such an outcome would be *bad*, because this person is not their true self, and that such satisfaction should not be the measure used to evaluate the outcome. This can be related to Berlin's distinction between an *empirical* and an *authentic* self. The former is the person as they appear today, while the latter is some *potential* version of this person [47,48]. The problem, then, is *who* should be considered when evaluating outcomes, the actual cared-for or some potential cared-for as interpreted by someone else.

Carers and relatives are also potential stakeholders worthy of consideration, and research shows that they tend to see therapy with dolls, for example, as 'demeaning, patronizing and inappropriate' [49,50]. However, studies show that other relevant stakeholders (children and young adults considering the use of robots in care, and specifically in the context of elderly family members), have relatively few qualms about using companion robots in care [7,51]. In a scoping review of research related to the use of PARO in care, Hung, et al. [2] argue that the literature contains a gap with regard to user (clinicians, policy-makers, families) and patient perspectives on the use of social robots in care.

4.3. Deception and the willing suspension of disbelief

Another concern about the use of social robots in care is *deception*. One possibility is quite simply that a person with dementia may not be able to distinguish a relationship with a robot from a relationship with a living person [21,36]. Another possibility is that the cared-for *knows* that a robot is not a living being, but still cannot help reacting to it *as if* it were [52].[3]

Deception is a term defined and used in various ways, and here I follow Williams [54] in seeing it as a) a breach of *trust* by the deceiver and b) *manipulation* by the deceiver of the deceived. However, we also need to identify the actors involved when discussing *robot* deception. I assume that a robot itself does not have agency and that it cannot be morally responsible for its actions. Someone else owns the actions of a robot, and it is this someone who deceives if a robot can actually be used for deception. I thus consider the perspective of the cared-for, and whether or not they are able to perceive the world around them accurately. The deceived and not the deceiver is thus in focus. The *intention* and even the *identity* of the deceiver is now of less importance.

A crucial point is that deception cannot be considered universally condemnable unless we simultaneously agree to abandon all sorts of games, theatre, movies, etc. Our societies are based in large part on the idea of benevolent or prosocial deception [55,56]. Still, we might decide that *robot* deception is not desirable or not desirable in the domain of care. However, studies show that robot deception is beneficial in terms of achieving higher levels of cooperation between humans and bots in games [57]. Traeger et al. [58] also show that robot deception can be used to improve *human* relationships and interaction by strategically making robots act 'vulnerable' and insecure.

While deception is mentioned here as a concern, several authors point out that social robot deception may in fact be beneficial and even a moral imperative if it leads to good *consequences* [59,60]. Meacham and Studley [31] note that *human* carers will also often be somewhat deceptive, and Isaac and Bridewell [61] note that robots *need* to deceive in order to be beneficial. They emphasise the fact that much deception is both benign and pro-social.

4.4. Slippery slope

A final argument against the use of social robots is based on the idea that this will lead to a situation in which robots displace human caretakers and eventually land us in a dystopia in which the elderly are taken care of by machines alone [24,37]. This is a slippery-slope argument that will not be assumed to be a logical necessity. I will consider it possible for social robots to be introduced into care only where they are beneficial, and agree with the idea that automation is as much about enabling humans to work more effectively as *replacing* human beings directly [38]. Concerns related to the displacement of humans by robots and the policy implications of this possibility will not be considered in detail here [62]. However, viewing social robots as *additions* to the care settings, and not replacements for humans, require us to provide both the resources and education necessary for the care sector to properly deploy and integrate social robots in the care setting [6,62].

4.5. The negative consequences summarised

Summing up, the negative aspects of deploying social robots in care involve all three aspects of *quality of care*. First, it may lead to structural changes with fewer human beings and more machines. This change, however, is only negative if we simultaneously consider *and* evaluate the nature of social robot care. Second, the *process* of being cared for by a robot is highly objectionable to those who view human-robot relations as either a) insufficient or b) harmful. Similarly, the process of social robot care may involve undesirable forms of deception. Some consider the process both demeaning and inhumane. Third, regarding *outcome*, the effects of human-robot relations could be that we are *diminished* [41]. If human contact is in fact a basic human need, the increased use of robots in care may lead to the deprivation of basic needs, and mental disintegration [43]. I will note that the *outcome*-related negative effects of social robots are largely hypothesised, based on anecdotal evidence, and somewhat vague [29].

5. The ethics of quality of care

From the preceding discussion, it follows that all aspects of quality of care, as derived from Donabedian [3], can be argued to be both positively *and* negatively affected by the use of social robots. Untangling this web of pros and cons requires the introduction of an ethical framework for assessing and weighing the various arguments—a task to which I shortly turn. Before doing so, however, I propose a modification of the model for evaluating *quality of care*. First, I propose that the *quality* factors must be evaluated *per person*, and that we separate *structure* from the other two, as this aspect is closely connected to the *quantity* of care and should be considered as such, also on a *societal* scale. This separation makes it easier to illustrate the balancing act that is care policy. The new model is shown in Fig. 2.

It is time to examine the core ethical concepts required to evaluate the various issues raised thus far. This examination will by necessity be brief, and the main idea is to highlight the various concerns that need to be accounted for in a policy for social robots in care and in particular to elucidate the trade-offs involved.

5.1. The ethics of outcomes

Beginning at the *end* of care, I examine how we might approach the evaluation of the outcomes of care. The outcomes, as we have seen, consist of both objective *and* subjective factors. It thus becomes clear that it is not simply a matter of weighing *process* against *outcomes*. We must also decide how to weigh the various forms of outcomes.

Sparrow [24] proposes that we focus on *objective* factors of care, as people's own evaluations of their situations may lead us astray. In particular, people have desires and preferences that may be objectively bad for them, which leads him to argue in favour of a sort of paternalism. This assumes that someone else could know better than I do what is good for me and thus have the right to override my preferences. This *might* be 'hard' paternalism, but it is also the basis of 'soft' paternalism, like *nudging* [63]. Both Sparrow [24] and Coeckelbergh [60] argue in favour of evaluating outcomes by considering the *capabilities* of the cared-for—an *objective* approach [64].

An alternative approach is to consider people's own judgements and preferences. This is the basic idea behind utilitarianism, which involves the quest for maximising the amount of *happiness*, when everyone's pleasure and pain is factored in [65]. Batayeh et al. [66] discuss *socially responsible innovation* in relation to healthcare, and state that this involves moving 'beyond tradition ideas of patient wellness that are only physiological'. While they propose that we include a range of *physical* outcomes, such as financial wellness and overall community wellness, I here propose that we could consider the patient's subjective perception of wellness.

One particular kind of utilitarianism is hedonism. It comes in two forms: psychological and ethical. The first is a theory of human *motivation*, and explains human behaviour as our desire for pleasure (and aversion to pain) [67]. The other form, ethical hedonism, is a theory of the evaluation of what is *good*, and it is this form of hedonism we might

[3] When the cared-for actually believes the robot is a living being, this can be called *full deception*. When the cared-for *knows* that it is interacting with a robot but still respond emotionally to the social signals emitted, this can be called *partial* deception [45]. Another set of categories for considering robot deception includes the distinction between *external state deception, superficial state deception,* and *hidden state deception* [53].

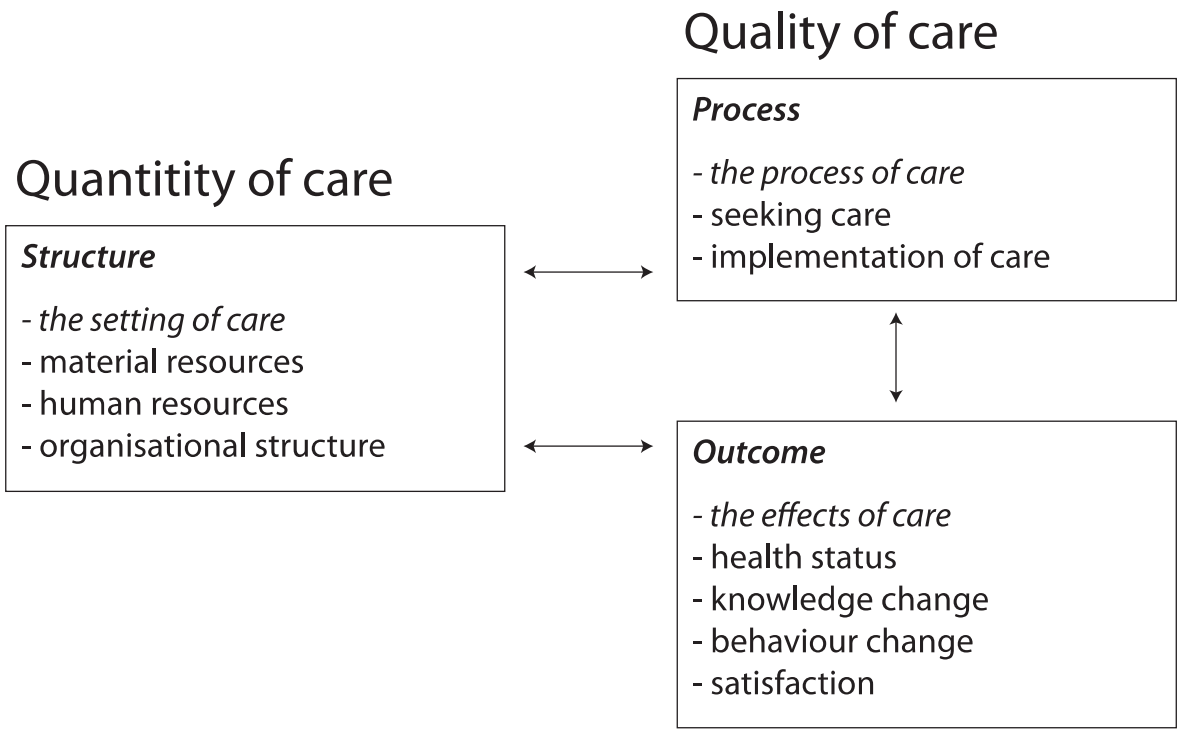

Fig. 2. Modified model of the quality of care.

apply if we want to emphasise the satisfaction of the cared-for, as measured by their happiness [68].

If we assume that social robots do provide pleasure, as experienced by the cared-for, this is considered *good* in a hedonistic ethic regardless of how this pleasure came about. However, when hedonism is coupled with utilitarianism, the opportunity to deny a person increased happiness arises if this increase comes at the cost of the *general* level of happiness.

We must, however, note that utilitarianism and hedonism are *related* but not identical to consequentialism. We could regard criteria other than individual happiness as the yardstick with which to evaluate various outcomes, such as the *capabilities* discussed by Nussbaum and Sen [64] or some combination of subjective and objective factors.

One problem, as pointed out by all the sceptics of deception and inauthentic relations with social robots, is that such an ethic would also accept some outcomes that many of us would consider to be undesirable. Sparrow and Sparrow [36], Sparrow [24], and Sætra [37] invoke the example of Robert Nozick's *experience machine.* This is a machine that could provide us with as many life-like experiences we would like while we are on life support in some tank [69]. While not actually performing any actions in a physical sense, we would *experience* life *as if* we did, with the added benefit, of course, that everyone could receive experiences tailored to maximise pleasure and joy, as scarce goods, wars, difficult human relationships, etc. would not come in the way of happiness [37].

The relevance of this example hinges on the assumption that deceiving the elderly with a robot is similar to the deception created by the experience machine. The *subjective* outcome may surpass anything achievable without such deception. The objective outcome, in terms of physiological health, for example, will be assumed to be at least equally good. Let us assume the elderly will be *less* stressed and *more* active due to their newfound happiness. The makers of an experience machine will certainly also be capable of providing physical stimulation in order to keep our bodies from withering while hooked up to the machine.

If satisfaction is increased and the objective outcome kept stable or even improved, what reason could we have to object to such a scenario? One answer is *process*. We could care about *how* the outcomes come about regardless of the actual outcomes, and this takes us to the ethics of *process*.

5.2. The ethics of process

How can the *process* lead us to say that an outcome that a person perceives as good is in fact bad? This is where we turn from consequentialist ethics focused on outcomes to *deontological* ethics of actions and process. For a consequentialist, the outcome can justify the means, but most deontologists will object to this.

Let us return to the example of the experience machine, which provides a good *subjective* outcome. Nozick [69] nevertheless states that being hooked up to such a machine is highly undesirable, akin to *suicide*.

The first reason for this is that we want to actually do things and not simply have the experience of doing them. This is similar to the objections based on the value of authenticity and negative effects of deception. Outcomes and experiences are themselves merely by-products of that which is valuable, which is an authentic process where we actually live our lives and interact with human beings. However, if the experiences and memories that were produced seem real, what possible reason could we have to care about how they came about?

Nozick's second reason is that we have a desire to 'be a certain way, to be a certain sort of person' [69]. If an elderly person were to be deceived by a robot seal into believing they were having a social relationship with a living being, for example, or that they were the captain of a pirate ship, and derived great joy from this, they would still not *actually* be the kind of person that they desired. But, again, why should we care, if the experience is real to the person in question? Sparrow [24] would say that *respect* would be lacking and that a person deceived in such a way had been deprived of *recognition.* Such a life would, according to Sharkey and Sharkey [21], not be *dignified.* People would be happy but they would not be treated as adult and competent beings—their happiness would be akin to the happiness produced in children when parents tell them that Santa is real.

Finally, Nozick [69] pre-empts the objections made by many opponents of the uncritical use of social robots, namely that the human-made experiences they produce are fundamentally different from *deeper* experiences. In order for this objection to be effective, however, we must

be able to point out exactly how human-made (robot) experiences are inferior to experiences produced by natural, or living, entities. We do not object to all experience created or assisted with human-made tools. If we are to stop social robots in care merely because they are human-made, we need more knowledge of just *how*, if at all, relating to them is harmful. Turkle [41] and Darling [70] provide compelling reasons for caution, but the harms of relating to robots must be better understood and studied if they are to inform policy for social robots in care.

If we are to disallow outcomes that are experienced as good, we could rely on what Nozick [69] calls *side constraints*. These are comparable to what is usually called *individual rights*, which is the most popular line of defence against unbridled utilitarianism. By appealing to such constraints or rights, I could stop you from optimising social happiness if that happiness was achieved by breaking some inviolable right. Such rights could, for example, be the right not to be killed, coerced, or defrauded. We could also add the right not to be *deceived* and the right to be treated with respect. Such rights could be inviolable and not amenable to the calculus of utilitarianism [73]. This is akin to the call for *Kantian* ethics in order to prevent the hedonism previously described [36]. This would imply that people should never be treated as a *means* to the happiness of others, but always as *ends*.

Unbridled hedonism and consequentialism might very well lead us to accept us all being hooked up to the experience machine, but inviolable rights not to be deceived or given due respect would also lead us into troubled water, for example by preventing hugely beneficial applications of new technologies because they infringe on some set rights [37].

6. Decision rules and robot policy

If we have a desire to create a society in which *technology* promotes a *good society*, we must determine how such a society treats individuals [5]. More than that, however, we must also allow for the fact that every society consists of *multiple* people, and any policy aimed solely at the individual runs the risk of having dire consequences for certain groups of individuals or even *all* individuals.

In general, it seems that social robots in care may have *negative* overall effects on the *process* of care. This is based on the objections which emphasise the beneficial nature of human contact for recognition, the need for respect, the value of what is natural (not man-made), and the general idea that deception is bad.

We have also seen that social robots can have beneficial effects on both objective and subjective health statuses. They may provide social interaction that is experienced as pleasant, which also stimulates sociability and stress reduction in those being cared for. We can imagine that use of social robots can lead to physiological and psychological benefits. However, we must not assume that such benefits would necessarily be maintained if robots completely *replaced* human beings in care. Let us assume that such a situation, as in Sparrow [24], would also lead to *negative* outcomes, caused for example by social needs not being met.

Finally, the question of how social robots affect the *structural dimension* of care is debatable. Sparrow [24] argues that care robots are *not* a panacea that will save the economy of care. However, we have assumed that a pure *social* robot without physical assistance capabilities can be quite cheap and that its use will lead to overall economic benefits if it allows for a lower ratio of carers to the cared-for.

Is care of a slightly lesser quality, provided for more people, a good thing? Our definition of *care* describes the meeting of *needs*, and I will assume that the needs in question are *non-trivial* and that not meeting said needs will have substantial negative consequences. A policy must define the point at which people should receive care. Providing care for *more* people than those who are in need is not desirable, which means that we have a target value for the quantity of care. When this target value is reached, we should do whatever we can in order to provide the best possible *quality* of care while maintaining the same *quantity*.

In the following I will make three assumptions. First, that the financing of the health sector is static. Second, that those who are defined as being in need of care have a *real* need for care, and we cannot redefine these needs in order to change the target quantity of care. Third, that the society in question does not have the financial possibility to provide human care for all in need of care. These assumptions are of course highly debateable, but they will allow us to see how the use of social robots can be evaluated.

First of all, social robots are potentially cost-effective, which means that they can be employed in order to change the structural conditions of care, allowing us to provide care for more individuals. Let us assume that this means that we increase the amount of material care resources available and slightly reduce the amount of human care workers.

If this is the *only* way to provide care for all those in need, the only real alternative in this scenario is to state that *no care* is better than *social robot care*. Since we see potential outcome benefits from the use of social robots, this statement will not be accepted.

After we have used social robots to provide care for all those in need, we must decide *how* best to deploy them and whether or not we should deploy *more* of them. If we assume that a social robot is not *better* than human caretakers, the main reason to introduce *more* social robots in care would be to rationalise the care sector and reduce the amount of money spent on care in general, but this is ruled out by the assumption of static financing. At this point, we are left to debate the relative benefits of human versus robot care.

We are, however, also at a point where the ethical principles involved in the use of social robots become clear. Employing social robots in care means that some who would otherwise receive human care, will receive care that is of slightly lesser quality. However, those who would otherwise *not* receive the care they need will receive much better care, even if that care is not optimal.

From a policy standpoint it makes little sense to refuse the use of social robots on the grounds that some individuals would be worse off than before. This would not be a case of using people as means for achieving the ends for others. If we consider the other alternative, *not* using social robots could mean that we used all those in need of care, who would now not receive care, as a means to an end for a lucky few who now received human care.

If we view policy formation as the choice among various scenarios for the future rather than the change of *actual* conditions today, we are in a much better place than if we have to consider the negative consequences for the *privileged actual people* of today. This is also where Rawls's *original position* and *veil of ignorance* become useful [71]. Without debating Rawls's premises and what is hidden by his veil, we can imagine the original position quite simply as a situation in which policy makers meet, without knowing who they themselves are, in order to make policy for the use of social robots in care based on the desiderata established thus far, as shown in simplified form in Table 1.

The outcome of such a process depends on an evaluation of the various benefit and drawbacks listed, and this evaluation is *necessarily* performed in light of some theory of ethics. It will often be an *implicit* system of ethics, but reaching an agreement on which policy is best will be very difficult unless we make the ethical assumptions we employ *explicit*.

The advantage of the thought experiment that is Rawls's original position is that alternative scenarios can be considered without considering the loss of *actual* persons. In such a situation it is obvious that we will not be able to make any reasonable decision without having some concern for the overall utility and consequences of our decision. Choosing a scenario where only the very best level of care is provided would *necessarily* lead to a lack of care for a vast number of persons, and there are no ethical systems that I know of that would suggest such a choice.

Finally, it is important to note that my focus on what Held [26] calls the 'dominant moral theories' and abstract reasoning is not universally accepted as a beneficial approach to the ethics of care or the ethics of healthcare robotics. Held [26] herself emphasises the claims of *particular*

Table 1
Policy desiderata for the use of social robot in care.

			Impact of social robots Positive: + Negative:
Quantity	Structure	Cost efficiency	+
		Speed	+
		Flexibility	+
		New opportunities	+
Quality	Process	Human contact	–
		Dignity	+ -
		Respect	–
		Recognition	–
		Speed and efficiency	+
	Outcome	Physical health	+ -
		Psychological health	+ -
		Experienced satisfaction	+
		Empowerment	+ -

others, and believes that abstract and universalistic rules will lead us astray. She further states that for 'most advocates of the ethics of care, the compelling moral claim of the particular other may be valid even when it conflicts with the requirement usually made by moral theories that moral judgments be universalizable, and this is of fundamental moral importance' [26]. Preston and Wickson [72] also emphasise how focusing on, for example, the feminist perspective and care ethics will lead to a different understanding of the impacts of technology than traditional approaches. This, they argue, is also compatible with a range of other ethical frameworks. Stahl and Coeckelbergh [39] similarly suggest that abstract philosophical approaches to the ethics of healthcare robotics is problematic, as these are too far from the *practical* aspects of innovation of design. I accept these objections to my approach, but emphasise that my focus on *policy formation* makes an abstract and philosophical approach to the ethical aspects of social robots in care necessary.

7. Conclusion

Social robots in care have the potential to result in both positive and negative consequences for care. These consequences have important implications for the ethical evaluation of the use of social robots and for the formation of a policy of social robots in care. I have presented Donabedian's [3] system of the three aspects of *quality of care*, which was modified to become a system for evaluating the quality *and* quantity of care. Structural aspects of care are seen as mainly related to the quantity of care while process and outcome are seen as core components of the quality of care. It is, of course, possible to state that *much* care is *quality* care, and the reader can easily make such a move without changing the gist of my argument.

The main potential benefits of social robots in care are related to structure (efficiency) and outcome. Social robots have the potential to increase care efficiency, thus enabling the provision of *more* care. They also have the potential to improve both physiological and psychological factors and the satisfaction of the cared-for.

On the other hand, the *process* of care is potentially negatively affected by the deployment of social robots, as they may reduce the amount of human contact. Robot care may also be undignified and disrespectful, and it involves a form of deception. In addition, relations with robots may be harmful, which is a factor of the *outcome* of care.

In evaluating these outcomes, I have made certain assumptions in order to highlight the most important trade-offs that must be made by policymakers. The major assumption is that the financing of the care sector is static. This implies that we cannot sidestep difficult choices by simply allocating more money to the sector in order to advocate against the deployment of social robots. Furthermore, the social robots I consider are potentially cost effective.

First, this means that one potential trade-off is some loss of quality in order to provide a larger quantity of care. The use of social robots may *decrease* the quality of care, but if this means that more people will receive care, this can easily be imagined to be an acceptable trade-off.

Secondly, policymakers must decide how to evaluate *process* versus *outcome*. The evidence suggests that social robots may provide outcome benefits at the cost of an objectionable process. Should policymakers decide that a good *process* and a strong focus on individual rights to respect, dignity, etc. is most important, they may have to accept worse outcomes.

Finally, a crucial question is whether or not to accept people's own evaluations of their situations or if we have a desire to overrule these in order to improve *objective* outcome gains.

Social robots in care thus have the potential to influence the quantity and quality of care, and policymakers should thoroughly examine the various aspects presented here. However, the policy desiderata in themselves are not sufficient for the formation of policy, as a system of ethics is a necessity for making the described trade-offs. This could be an *implicit* and unacknowledged system of ethics. While it *could,* it seems clear that making this system of ethics explicit will make the process more transparent, and by consequence possibly more legitimate.

Author statement

Henrik Skaug Sætra: All parts of the article.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.techsoc.2020.101383.

References

[1] L. Bodenhagen, S.-D. Suvei, W.K. Juel, E. Brander, N. Krüger, Robot technology for future welfare: meeting upcoming societal challenges–an outlook with offset in the development in Scandinavia, Health Technol. 9 (3) (2019) 197–218.
[2] L. Hung, et al., The benefits of and barriers to using a social robot PARO in care settings: a scoping review, BMC Geriatr. 19 (1) (2019) 232, https://doi.org/10.1186/s12877-019-1244-6.
[3] A. Donabedian, The quality of care: how can it be assessed? Jama 260 (12) (1988) 1743–1748.
[4] M. Coeckelbergh, Technology and the good society: a polemical essay on social ontology, political principles, and responsibility for technology, Technol. Soc. 52 (2018) 4–9.
[5] C. Griffy-Brown, B.D. Earp, O. Rosas, Technology and the good society, Technol. Soc. 52 (2018) 1–3.
[6] G. Toms, F. Verity, A. Orrell, Social care technologies for older people: evidence for instigating a broader and more inclusive dialogue, Technol. Soc. 58 (2019) 101111.
[7] H.L. Bradwell, R. Winnington, S. Thill, R.B. Jones, Ethical perceptions towards real-world use of companion robots with older people and people with dementia: survey opinions among younger adults, BMC Geriatr. 20 (1) (2020) 1–10, https://doi.org/10.1186/s12877-020-01641-5.
[8] A. Poulsen, O.K. Burmeister, Overcoming carer shortages with care robots: dynamic value trade-offs in run-time, Australasian Journal of Information Systems 23 (2019).
[9] T. Kochetkova, An overview of machine medical ethics. Machine Medical Ethics, Springer, 2015, pp. 3–15.
[10] T. Fong, I. Nourbakhsh, K. Dautenhahn, A survey of socially interactive robots, Robot. Autonom. Syst. 42 (3–4) (2003) 143–166.
[11] K. Dautenhahn, Getting to know each other—artificial social intelligence for autonomous robots, Robot. Autonom. Syst. 16 (2–4) (1995) 333–356.
[12] C. Breazeal, Toward sociable robots, Robot. Autonom. Syst. 42 (3–4) (2003) 167–175.
[13] D. Feil-Seifer, M.J. Mataric, Defining socially assistive robotics. 9th International Conference on Rehabilitation Robotics, ICORR 2005., 2005, pp. 465–468, 2005: IEEE.
[14] K. Obayashi, N. Kodate, S. Masuyama, Can Connected Technologies Improve Sleep Quality and Safety of Older Adults and Care-Givers? an Evaluation Study of Sleep Monitors and Communicative Robots at a Residential Care Home in Japan, Technology in Society, 2020, p. 101318.
[15] A. Tapus, C. Tapus, M.J. Mataric, The use of socially assistive robots in the design of intelligent cognitive therapies for people with dementia. IEEE International Conference on Rehabilitation Robotics, IEEE, 2009, pp. 924–929, 2009.
[16] R.-M. Johansson-Pajala, et al., Care robot orientation: what, who and how? Potential users' perceptions, International Journal of Social Robotics (2020) 1–15.

[17] Blue frog robotics. "Meet Buddy." http://www.bluefrogrobotics.com/robot/(accessed.
[18] Paro Robots, PARO therapeutic robot (accessed, http://www.parorobots.com.
[19] LEKA, Help exceptional children live exceptional lives (accessed, https://leka.io.
[20] Cutii, Cutii (accessed, https://www.cutii.io/en/.
[21] N. Sharkey, A. Sharkey, The eldercare factory, Gerontology 58 (3) (2012) 282–288.
[22] G. Marcus, E. Davis, Rebooting AI: Building Artificial Intelligence We Can Trust, 2019. Pantheon.
[23] M. Coeckelbergh, 'How I learned to love the robot': capabilities, information technologies, and elderly care. The Capability Approach, Technology and Design, Springer, 2012, pp. 77–86.
[24] R. Sparrow, Robots in aged care: a dystopian future? AI Soc. 31 (4) (2016) 445–454.
[25] R. Bemelmans, G.J. Gelderblom, P. Jonker, L. De Witte, Socially assistive robots in elderly care: a systematic review into effects and effectiveness, J. Am. Med. Dir. Assoc. 13 (2) (2012) 114–120.
[26] V. Held, The Ethics of Care: Personal, Political, and Global, Oxford University Press., 2006.
[27] D.E. Bubeck, Care, Gender, and Justice, Oxford University Press, Oxford, 1995.
[28] P. Bowden, Caring: Gender-Sensitive Ethics, Routledge, London, 1997.
[29] D.J. Gunkel, Robot rights, MIT Press, London, 2018.
[30] E. Levinas, P. Nemo, Ethics and infinity, CrossCurrents 34 (2) (1984) 191–203.
[31] D. Meacham, M. Studley, Could a robot care? It's all in the movement, in: P. Lin, J. Ryan, K. Abney (Eds.), Robot Ethics 2.0, Oxford University Press, 2017.
[32] V. Bernabei, et al., Animal-assisted interventions for elderly patients affected by dementia or psychiatric disorders: a review, J. Psychiatr. Res. 47 (6) (2013) 762–773.
[33] W. Moyle, et al., Exploring the effect of companion robots on emotional expression in older adults with dementia: a pilot randomized controlled trial, J. Gerontol. Nurs. 39 (5) (2013) 46–53, https://doi.org/10.3928/00989134-20130313-03.
[34] N. Sharkey, A. Sharkey, Granny and the robots: ethical issues in robot care for the elderly, Ethics Inf. Technol. 14 (1) (2012) 27–40.
[35] T. Vandemeulebroucke, B. Dierckx de Casterlé, L. Welbergen, M. Massart, C. Gastmans, 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, 2019.
[36] R. Sparrow, L. Sparrow, The hands of machines? The future of aged care, Minds Mach. 16 (2) (2006) 141–161, https://doi.org/10.1007/s11023-006-9030-6.
[37] H.S. Sætra, First, they came for the old and demented, Human Arenas (2020) 1–19, https://doi.org/10.1007/s42087-020-00125-7.
[38] V.C. Müller, Ethics of artificial intelligence and robotics, in: E.N. Zalta (Ed.), Stanford Encyclopedia of Philosophy (Summer 2020 Edition), 2020.
[39] B.C. Stahl, M. Coeckelbergh, Ethics of healthcare robotics: towards responsible research and innovation, Robot. Autonom. Syst. 86 (2016) 152–161.
[40] S.L. Anderson, M. Anderson, Towards a principle-based healthcare agent. Machine Medical Ethics, Springer, 2015, pp. 67–77.
[41] S. Turkle, Alone Together: Why We Expect More from Technology and Less from Each Other, 2017. Hachette UK.
[42] A.H. Maslow, Motivation and Personality, Pearson Education, Delhi, 1987.
[43] E. Fromm, Escape from Freedom, Holt, New York, 1994.
[44] T. Yıldız, Human-computer interaction problem in learning: could the key Be hidden somewhere between social interaction and development of tools? Integr. Psychol. Behav. Sci. 53 (3) (2019) 541–557.
[45] H.S. Sætra, The parasitic nature of social AI: sharing minds with the mindless, Integr. Psychol. Behav. Sci. (2020) 1–19.
[46] B. Reisberg, E.H. Franssen, L.E. Souren, S.R. Auer, I. Akram, S. Kenowsky, Evidence and mechanisms of retrogenesis in Alzheimer's and other dementias: management and treatment import, Am. J. Alzheimer's Dis. Other Dementias 17 (4) (2002) 202–212.
[47] I. Berlin, Liberty, Oxford University Press, Oxford, 2002.
[48] I. Carter, A Measure of Freedom, Oxford University Press, 1999.
[49] I.A. James, L. Mackenzie, E. Mukaetova-Ladinska, Doll use in care homes for people with dementia, Int. J. Geriatr. Psychiatr.: A journal of the psychiatry of late life and allied sciences 21 (11) (2006) 1093–1098.
[50] L. Mackenzie, I.A. James, R. Morse, E. Mukaetova-Ladinska, F.K. Reichelt, A pilot study on the use of dolls for people with dementia, Age Ageing 35 (4) (2006) 441–444.
[51] R.A. Søraa, P.S. Nyvoll, K.B. Grønvik, A. Serrano, Children's perceptions of social robots: a study of the robots Pepper, AV1 and Tessa at Norwegian research fairs, AI & society, 2020, https://doi.org/10.1007/s00146-020-00998-w.
[52] C. Breazeal, Designing Sociable Robots, MIT press, Cambridge, 2004.
[53] J. Danaher, Robot Betrayal: a guide to the ethics of robotic deception, Ethics and Information Technology, 2020, pp. 1–12, https://doi.org/10.1007/s10676-019-09520-3.
[54] B.A.O. Williams, Truth & Truthfulness: an Essay in Genealogy, Princeton University Press, Princeton, 2002.
[55] M.L. Knapp, M.S. McGlone, D.L. Griffin, B. Earnest, Lying and Deception in Human Interaction, Kendall Hunt Publishing, 2015.
[56] E.E. Levine, M.E. Schweitzer, Prosocial lies: when deception breeds trust, Organ. Behav. Hum. Decis. Process. 126 (2015) 88–106.
[57] F. Ishowo-Oloko, J.-F. Bonnefon, Z. Soroye, J. Crandall, I. Rahwan, T. Rahwan, Behavioural evidence for a transparency–efficiency tradeoff in human–machine cooperation, Nature Machine Intelligence 1 (11) (2019) 517–521.
[58] M.L. Traeger, S.S. Sebo, M. Jung, B. Scassellati, N.A. Christakis, Vulnerable robots positively shape human conversational dynamics in a human–robot team, Proc. Natl. Acad. Sci. Unit. States Am. 117 (12) (2020) 6370–6375.
[59] A. Sharkey, N. Wood, The Paro seal robot: demeaning or enabling, Proceedings of AISB 36 (2014).
[60] M. Coeckelbergh, Care robots and the future of ICT-mediated elderly care: a response to doom scenarios, AI Soc. 31 (4) (2016) 455–462.
[61] A.M.C. Isaac, W. Bridewell, White lies on silver tongues: why robots need to deceive (and how), in: P. Lin, J. Ryan, K. Abney (Eds.), Robot Ethics 2.0, Oxford University Press, 2017.
[62] R.U. Ayres, S.M. Miller, Robotics and conservation of human resources, Technol. Soc. 4 (3) (1982) 181–197.
[63] C.R. Sunstein, Why Nudge?: the Politics of Libertarian Paternalism, Yale University Press, 2014.
[64] M. Nussbaum, A. Sen, The Quality of Life, Oxford University Press, 1993.
[65] J.S. Mill, J. Bentham, Utilitarianism and Other Essays, 1987. Penguin UK.
[66] B.G. Batayeh, G.H. Artzberger, L.D. Williams, Socially responsible innovation in health care: cycles of actualization, Technol. Soc. 53 (2018) 14–22.
[67] S.M. Cahn, C. Vitrano, Happiness and Goodness: Philosophical Reflections on Living Well, Columbia University Press, 2015.
[68] R. Audi, The Cambridge Dictionary of Philosophy, Cambridge university press, Cambridge, 1999.
[69] R. Nozick, Anarchy, State, and Utopia, Basic Books, New York, 1974.
[70] K. Darling, Who's Johnny?'Anthropomorphic framing in human-robot interaction, integration, and policy, in: G.B.P. Lin, K. Abney, R. Jenkins (Eds.), ROBOT ETHICS 2.0, Oxford University Press, 2016.
[71] J. Rawls, A Theory of Justice, Harvard university press, 2009.
[72] C.J. Preston, F. Wickson, Broadening the lens for the governance of emerging technologies: care ethics and agricultural biotechnology, Technol. Soc. 45 (2016) 48–57.
[73] Thomas Nagel, Foreword by Thomas Nagel, in: R. Nozick (Ed.), *Anarchy, state, and utopia*, Basic Books, New York, 2013.