Alice Does not Care
Or: Why it Matters That Robots "Don't Give a Damn"

Imke von Maur

1 Introduction

In his documentary *Alice cares* (2015), Sander Burger presents a pilot project of researchers in Amsterdam in which a robot called Alice is introduced as a new friend to three elderly women. In this text, I will raise serious doubts that Alice, as the title of the movie claims, cares. On the contrary, I point out why Alice does not and in principle cannot care. This circumstance is illustrative of John Haugeland's utterance with regard to the general problem of artificial intelligence (AI) – namely that computers "don't give a damn". All that robots do is pattern recognition and giving a seemingly adequate output to a given input. If the capacity to discriminate between *this* and actual care gets lost, as I argue, there is the danger that not only the concept of *care* changes, but ultimately the practice of caring.

Alice is a paradigmatic example of this and highlights, on the one hand, the structural presuppositions that lead to the development of such technology in the first place, and, on the other hand, the severe consequences this has for society and our understanding of what it means to be human and to live a good life. Taking seriously the quintessence of this text comes with consequences for the research and implementation of so-called social robots, which are supposed to care not only for elderly people but be companions of lonely teenagers and adults in times of a pandemic.

I will argue that implementing robots in order to care and to reduce loneliness not only fails – for robots simply cannot care – but has the potential to make things worse in a systematic way. An ethical assessment of care robots thus should not only tackle the concrete outcomes of a care robot in a given scenario – I call this a *functionalist-individualist approach* – but needs to be concerned with the broader socio-political structural presuppositions and effects.

One *structural effect* of normalizing the implementation of care robots could be that humans will not only stay lonely while interacting with robots, but that they will give up expectations of real care and true relationships in the first place. This might lead to a severe change in (expectations within and about) social interactions at a large scale. The effects a continuous interaction with a care robot might have on the affective and social repertoire of a person thus exceeds the scope of this individual. An example for a *presupposition* of normalizing simulated care is the techno-solutionist narrative, according to which societal problems can be satisfactorily dealt with by implementing technological solutions. This narrative makes intelligible the implementation of care robots to researchers, tech-companies, the public, political decision-makers and ultimately the customer in the first place. There are other potential imaginations and narratives on how to provide care which should be considered in an ethical assessment of care robots.

The aim of this paper is twofold: Firstly, I demonstrate why care robots do not and in principle cannot care. This is a rather analytical point. Connecting to this, I make the normative argument that it in fact matters that humans *actually* care about another instead of *simulating* to do so. This normative step provides the ground for my second aim in this paper, namely to draw attention to the necessity of a larger socio-political approach for an ethics of care robots (see also Coeckelbergh 2022 who recently argues for a "political philosophy of AI"). That is, I aim to provide the grounds for a different focus for assessing the appropriateness of using care robots to reduce loneliness which goes beyond local-individualistic and functionalist perspectives but considers the broader socio-political horizon. It is this structural perspective I am concerned with throughout the text. Thus, it is important to clarify that I am not arguing at the level of concrete individuals interacting with robots. The "ethical subject" I address in this text is not a concrete subject but society at large in its responsibility for how discourses are shaped and for which narratives become powerful. The caveat being made here thus concerns the temptation to take what I argue for as prescriptive suggestions for prohibitions on the individual level. That is not the goal. The goal is to raise awareness to the potential structural consequences of the wrong assumption that robots could care about us. That is, consequences for societally shared narratives and imaginations about how to live a good life (together) and on related practices and institutionalizations, among others.

2 Alice – "The care of tomorrow"?

Alice: What makes you happy?
Ms. Remkes: I haven't figured out yet. [pause] "What makes you happy?" [laughs seemingly ashamed and looks away] I have to think about that.
Alice: What else could I do for you?
Ms. Remkes: [face starts to get angry] I don't feel like having a robot in my home. [looks around with a dissatisfied face] I prefer a living human being.
Alice: Oh, that's a shame. Thank you for the conversation. Maybe I'll see you soon.
Ms. Remkes: We'll wait and see.

The first version of Alice that is used in the documentary is a small doll with a puppet-like female face that should feel soft and skin-like. The body, though, is that of a plastic toy and does not resemble a human, although there is a torso, two arms and two legs. Alice is small, it can only sit on the couch or a chair like a doll but not move itself or walk around, let alone do any physical task for those it is supposed to care about. The 'caring' is meant to be one of companionship: "Alice's first mission is to assume a social role and be treated as a social entity. That is why we designed a social robot that is both, a friend as well as a connector to others. (Alice Cares Promo 2021)" On the webpage "Alice cares", its main selling point is announced as its ability to "decrease feelings of loneliness": By engaging in "a conversation, see[ing] what her companion is doing and by scanning her environment, Alice is able to react and adapt to situations in her surroundings." (ibid.) The robot is furthermore said to remember things its counterpart said, to ask them questions, and to motivate them to engage in specific activities like going for a walk, singing a song or writing a letter.

The documentary encourages the audience to be skeptical at the beginning of the story since the three women are initially not portrayed to be enthusiastic about the robot. But, over the course of the story, the subjects seem to become more and more attuned to the robot, to accept it and even to like it, to engage in conversations with it and ultimately find a companion in it. I interpret this narrative of a *seemingly* stepwise acceptance to be manipulating the viewers of the documentary and think it obscures a serious engagement

with both the problem and the three women.[1] One scene at the end of the documentary shows how Ms. Remkes goes to a café with Alice (the scientist carries Alice to this place for it cannot walk). Ms. Remkes seems to be proud of being there with Alice and talks to the waitress about her being part of the research and pilot project. This scene is meant to elicit the feeling in the viewer that Ms. Remkes is happy because of Alice being her companion. But this is a typical instance of the "Hawthorne-effect": The circumstance that she is part of a study changes her feelings and behavior.[2] Ms. Remkes gets attention, people are aware of her because of the camera team, of the robot and the special situation. She is *recognized* by other *humans* as being special, important, ultimately: *as being there*. Another misleading thought the documentary suggests is that, even though Alice might not completely compensate for human interaction, it is at least sophisticated entertainment, like an advanced TV, one that supposedly reacts and responds instead of being merely looked at. There seems to be an intuitive suggestion to say that this is better than nothing. But nothing is not the only alternative. The problem is already framed in a way in which it seems to be the only available and thus inevitable solution to rely on technological progress in order to counter societal problems and developments. The *crucial* question, that motivates the following considerations is: How is it that researchers, political decision-makers and the broader public really consider it possible for a machine to care?

In the following section, I shed light on the complexity and depth of the phenomenon of care that I take to be indispensable for humans and that (not

1 I will sketch some indicators for this interpretation in what follows without the aim to generalize this critique by inferring that *all* documentaries about care robots *necessarily* have to work this way. Also, the manipulative character of the documentary alone does not prove that the elderly women portrayed or even any elderly person engaging with a care robot has to be unhappy with that. What I aim at with this opening here is to highlight the framing of one contribution for a discourse about a normative assessment of care robots – namely this specific documentary – which I find problematically suggestive.

2 From 1927 to 1932 researchers conducted a field study in the Hawthorne department of the American Western Electric company to figure out, which factors would increase the productivity of the workers (Roethlisberger & Dickson 1939). What is now known as the "Hawthorne effect" is the influence of the very setting of a study on the attitude and behavior of the subjects under investigation (Sanders & Kianty 2006: 59ff.). This phenomenon, among other similar effects, is also investigated under the term "reactivity", for instance in a recent and encompassing research project by Marion Godman, Caterina Marchionni and Julie Zahle (Reactivity Project 2021) .

only) individuals in the case study of Alice miss. Afterwards, I reveal the understanding of care underlying care robots and show that the two have nothing to do with each other, i.e., that Alice does not care.[3]

3 Caring: Being (taken as) a person

Care is a complex and manifold phenomenon. To care means, first and foremost, to not be indifferent to what happens to another. This is not bound to the necessity that the other is of high personal import to me like friends or beloved ones, not even that a teacher or nurse for instance has to like all their students or patients wholeheartedly. But it manifests differently in caring bodies what happens to the one they *care about*. For the topic of the paper, I am concerned with *caring about* as "a mental capacity or a subjective state of concern" (van Wynsberghe, 2013: 414) and not *caring for* meant as "an activity for safeguarding the interests of the patient" (ibid.). While I believe the latter is not properly delivered without the attitude being at issue in the former one, an argument for this is not the topic of debate here. As I want to provide the ground for an argument against emotional bonds with robots designed to be social *companions* – it cannot be meant that "care" is only the performance of tasks (*care for*) like carrying patients to bed or reminding them to take their medicine, but must include an *attitude*, by which a companion is characterized (*care about*).[4] It is not simply "caregiving" in the sense of sustenance but care in the sense of a meaningful relationship I am concerned with in the following.[5]

---

3 In this paper, I am concerned with robots designed to be companions and to reduce loneliness. There are artificial systems called robots or care assistants as the famous seal Paro used for patients with severe dementia, which might work as sophisticated entertainment but are not intended to be a companion who cares. What I discuss in this paper is not about Paro or robot dogs or other such devices.

4 "In the field of care ethics, Joan Tronto claims that good care is the result of both a caring attitude in combination with a caring activity (Tronto 1993). In other words, a marriage between the dimensions of caring about and caring for." (van Wynsberghe 2013: 414)

5 I will solely be concerned with human-human relationships in this section and will remain silent about the possibilities of care-relationships between humans and animals or plants.

In this sense, as said, a caring person is *affectively not indifferent* to what is going on. This is self-explanatory for intimate relations characterized by a strong emotional bond – loving relations, friendships, family relationships etc. But also, in a relationship of care in the medical or educational context the involved subjects react affectively, depending on the other's situation. This becomes especially clear when concrete encounters are considered. In situations where humans engage with each other face-to-face, they look at each other, recognize vocalizations and facial expressions and their meaning in this specific context against the background of a *shared space of meaning*. Humans and their encounters are inevitably situated in concrete practice-specific contexts, in which something is "at issue and at stake" (Rouse 2002) that is often hard to put into words. A person is not just sad. A concrete human being with concrete needs, desires and hopes, with a concrete and specific (affective) biography makes sense of their situation (affectively).[6]

To care about this person means to bring them into existence not only *as a person* (rather than an object) but as *this very concrete individual*.[7] A caring person brings into existence the other one as a valuable being, a concrete individual with specific needs, concerns, beliefs, values and so on. Phenomeno-

6 Elsewhere I elaborate on the human capacity to (tacitly) navigate in such spaces by adopting a practice theoretical and phenomenological approach to the disclosure of meaning, which is not only characterized by affectivity but by sociality and the knowledge about relevant rules of practice specific "games", as Bourdieu would calls this (von Maur 2018 and 2021). The tacit (social) knowledge humans incorporate goes beyond what is implementable in a machine, inter alia because there are too many and too subtle things to be known in order to skillfully engage in such domain specifically meaningful "little worlds" (ibid.).

7 The claim is not meant in the sense that the human being would not exist without this act of recognition. For sure, the flesh of a creature does not fade away if nobody recognizes it as a person. But as Judith Butler claims with regards to the question the life of whom is grievable, the ontological, epistemological and ethical aspects of "apprehending a life" are inextricably intertwined (2009: 2ff.). If persons are not apprehended as persons, not *brought into existence* as vulnerable beings, they do not appear in our repertoire of the ones we grieve for. Butler is especially interested in the framings, for instance of media, which determine the life of whom is apprehended and of whom not. I take it to be worth further research to apply her framework – which in her book is applied to war and its victims – to the field of the elderly and children. These societal groups are especially vulnerable and deserve a cautious consideration and treatment. Yet, often they do not seem to be apprehended in a way necessary to initiate appropriate conditions for care and education. Children and the elderly rather seem to be effortful and obstacles for the working subject in a neoliberal society.

logically this is nicely illustrated by philosophers like Jean-Paul Sartre or Emmanuel Levinas who emphasize the import of concrete encounters for ethics, i.e. by pointing out what it means to look our counterpart into their "Antlitz" (Levinas). This is an act of personification, an act where we recognize the other one as another ethical subject, a subject with unimpeachable human dignity and human rights. Martin Buber (1973) speaks of "making the other one present" [*personale Vergegenwärtigung*] – meaning to perceive the other one as a whole and as being unique – and already 50 years ago criticized that in modern times there is an omnipresent analytical and reductive perception of others rather than such a perception of the other one as a concrete valuable being (1973: 285). If a person is not able to recognize the other person in the way just described, if they are not able to be affected, it is hard to see how they can care. To care about someone thus means to engage in a relationship and to make the other person matter, to decide that their concerns and well-being matter to oneself – ultimately: to make the concerns of another person one's own concerns.

This does not imply that to care means to sacrifice. There is no universal or personal duty involved in caring. Instead, to care presupposes that the caring person *decides autonomically* to care about the other, to take part in their life.[8] Although this is quite demanding, people feel existentially threatened if they imagine that *nobody* cares about them *in this way*. For sure, the degree in a loving relationship or intimate friendship is much higher when it comes to this feature of relationships, but, still, this is what makes up other caring relationships as well – i.e., educational or medical ones. Crucially, in any caring relationship one *responds* to the other as well as taking *responsibility* for them.

Although taking the counterpart seriously without it being a sacrifice or involving patronizing attitudes is key for caring relations, there are certainly differences in the kind of (a)symmetry of caring relations. To care about children who have not developed a decisive degree of autonomy yet, or to care about an elderly person who has lost this autonomy (to a certain degree), are

8 This can be nicely illustrated by help of Tzvetan Todorov's (1993 [1991]: 80-101) distinction between care and mercy (and solidarity or charity). The merciful person sacrifices and assumes a burden, whereas the caring person just cares. This does not mean that this is not effortful or experienced painfully. I cannot but care about my children although this is often demanding and painful, and in the same manner, real care in the medical or educational or other contexts, implies the attitude of making the problems of the other one's own – not in the intensity of a parent-infant relation, but in a way that goes beyond mercy.

instances of rather asymmetrical relations. The one who cares is the one who takes the responsibility for the less autonomous member of the relationship. Yet, it is key to *take seriously* the care-receiving person *as a person* and to take seriously that there is, also for a less autonomous person, like a child or elderly person, a possibility to "care back".[9] Thus, also the one who cares is a concrete counterpart, an individual speaking with their own voice, one who can also disagree with the other, be angry at them or be in need of consolation. If a neighbor, relative or nurse cares for and about an elderly person, the care-giver plays a role in their life. At least in such a relationship of care that lasts for a longer period, the elderly individual also includes the care-giver in their thoughts and considerations, takes an interest in their well-being, achievements or sufferings. Phenomenologically, in a concrete situation there are two parts who can establish *resonance* (in the sense of Hartmut Rosa 2016). That is, there is not a mere echo of one's own voice but a being who can understand, misunderstand, like or dislike me, who can be similar in some regards rather than others and the engagement with whom can feel comfortable – I can reach the other one in Rosa's sense – or it feels alienating because I feel misunderstood. Both parts in such a relationship that is phenomenologically experienced in specific ways need to be taken as fallible, but first and foremost unconditionally valuable individuals – independently of whether the relationship is (always) experienced as harmonic or rather discordantly.

It is important to keep in mind that I am not claiming that any human professional care-giver could or should care about any of their patients at any time in an emotionally exhaustive way. A certain degree of "professional" (affective) distancing seems to be crucial for professional care-givers in order to stay healthy and not to exploit their (affective) resources. The characterization of care just sketched is meant to highlight crucial aspects of the *attitude* of caring about another *in general* – as a state of *concern* that is characterized by not being indifferent. To care (in the sense of the German word "Sorge" rather than "Pflege") is an affective attitude and practice that provides the basis or

9 I thank the Research seminar of the Centre for Ethics in Pardubice for discussing this point with me. The idea that to be *cared about* implies the possibility to *care back* makes clear that not to be lonely also means that one is able to partake in the life of another one – namely to worry about their well-being as well. Yet, this is not possible for *any* person who still might be truly cared about – think for instance about patients with severe dementia or people in vigil coma. I do not claim that one cannot truly care about these people who cannot care back in the sense sketched.

background of a relationship. A professional care-giver can care in this sense without necessarily being highly emotionally involved in any interaction with their patients. The crucial aspect rather is the real interest they take in the well-being of the other one, that it makes a difference to them whether the patient is happy or sad, is in a good or bad bodily condition, suffers from social isolation or shares parts of their life with others.

Before moving on to explain why a robot does and cannot care in the way described here, I illustrate the concept of care that in fact *is* at issue in so-called care or social robots – i.e., an atomistic approach to care that reduces the complex meaningful phenomenon just described, to input-output-processes in a functionalistic manner. This 'care' implemented in robots has nothing to do with *caring about*, as I described it. Yet, normalizing the use of the concept 'care' for denoting simulated care involves the danger of powerfully disempowering narratives and imaginations about what it means to entertain a relationship, what it means to care and to be cared about and ultimately, what it means to live a good life.

4 Reductionistic encounters: A functionalist understanding of care

"The trouble with artificial intelligence is that computers don't give a damn". This utterance of John Haugeland (1998: 47) highlights the thesis that robots cannot care in the sense sketched above. Underlying the incapacity to "give a damn" is the robots' incapacity to be affected. I take for granted in this paper that robots do not have emotions, that they lack consciousness and thus any kind of first personal experience in terms of "what it is like" (Jackson 1982).[10] The difference between so-called strong and weak AI (Searle 1980) does not only apply for cognitive but also for affective processes. The thesis of weak AI, namely that artificial systems are able to *recognize* and *simulate* human capabilities in ways that we are unable to distinguish them from human comportment, thus concerns *emotion recognition* and *emotion simulation* in the context

10 From an epistemological point, being more precise would be to say that we do not know. We can never know if a robot has feelings, as we cannot know if another person or an animal has feelings. But we have more than good reasons to believe that animals are able to suffer and that our fellow humans are conscious, while these reasons are not present for an alleged consciousness and affectivity of machines.

of this chapter. The relevant question I aim to critically assess here thus is not whether robots do really feel, but rather if they recognize and simulate feelings sufficiently well to make the humans interacting with them believe they really do have emotions and understand theirs.[11] Yet, it might not even be necessary that the interacting human believes that the robot actually experiences emotions and understands the person. The thesis can also be that having the *impression* that the robot recognizes and experiences emotions (while being aware that it does not) might just fulfill the function that an interaction can develop in a more or less expected manner – and this, in the case of humans, includes emotions to be recognized and displayed by both interactants.

The key assumption which is relevant for my present purpose is that robots understand and adequately react to their counterparts' emotions and *by this* care. In doing so, that is the idea I aim to argue against, they engage in conversations, console or motivate the cared-about person – the types of things typically done by the social companions or friends they are supposed to be instantiations of (see again the Alice Cares Promo 2021, quoted in the introduction already). This is a reductionist and functionalist, behavioristic take on what it means to care that does not allow us to capture the complex phenomenon at issue. More troublesome so, the development of machines that should care presupposes as well as leads to a reductionist concept not only of care, but also implies a reductionist concept of emotions and of social relationships. To make explicit what I do (not) mean with this: When saying *reductionistic,* this refers firstly to the concept and practice of care. It is a concept and practice of care being *reduced* to input-output-relations independent of their meaning that I argue against. This does not imply any overall argument against functionalist approaches within the philosophy of mind or elsewhere. What I argue against is the functionalist-individualist ethical approach for assessing the employment of care robots which is both, presupposing and resulting in a reductionist concept of care. In the following I demonstrate the reductionistic functionalist concept of care that I aim to criticize by taking

11 The thesis of strong AI, on the contrary, amounts to the claim that artificial systems not only simulate but actually exhibit what is called human intelligence, i.e. in our context: emotions. As I presuppose that computers as of today are not able to be affected, this thesis will not be considered further. In how far the thesis of weak AI holds is under debate since the infamous *Turing test* as well Joseph Weizenbaum's ELIZA, which are supposed to test whether a human could identify if they are interacting with a machine or a human in a conversation on screen.

a look at how emotion recognition and simulation works in so-called care or social robots by considering emotion recognition via facial and voice recognition.[12]

Emotion recognition software typically operates on the assumption that there are universal and basic facial expressions for distinct emotion types. These assumptions go back to the work of psychologist Paul Ekman and colleagues (1978, 2003), as does the thesis that there are six basic emotions (joy, surprise, anger, sadness, disgust, and fear), each of which is accompanied by prototypical facial expressions claimed to occur in and be understandable by all humans. Ekman and his colleagues developed the so-called "Facial Action Coding System" (FACS) which analyzes facial expressions by devoting so-called "Action Units" (AU) to the observable facial movements. This leads to a classification of facial expressions being representable as code. Any basic emotion is describable as a combination of characteristic action units. Joy for instance is describable as a combination of raising cheeks and raising corners of the mouth. Furthermore, the intensity of an emotion can be graded on a scale and also head- and eye movements and typical behavioral patterns can be added to the analysis (cf. also Misselhorn 2021). Artificial emotion recognition relying on facial expression recognition proceeds in three steps (ibid.: 22): First, the face is recognized *as a face* (as an object being different from a chair, a window, etc.), second, the features described above are extracted, and third, the emotion is classified with regards to the combination of extracted features.

This procedure is a paradigmatic example of what Winograd & Flores called the "rationalistic tradition" in the sciences (1986: 14), namely a specific way in which science frames and addresses its questions. On their account, scientific research in this tradition "consists of setting up situations in which observable activity will be determined in a clear way by a small number of variables that can be systematically manipulated" (1986: 16). Accordingly, the approach to a research question in this spirit is to find identifiable objects and interaction rules which, in combination, provide an answer to the initial question (1986: 15). Such an approach presupposes an *atomistic assumption*, as Schuetze and von Maur (2021: 8) call it, namely that "the world can be split into parts and then be analyzed in terms of these building blocks as well as

12 There are also other methods by which artificial systems should be enabled to recognize emotions, like sentimental analysis or using biosensors (see Misselhorn 2021, chapter 2).

the rules according to which they interact". While such an approach might be helpful for some research questions, when used for a complex phenomenon such as "human affectivity" or "care" it neglects other essential features making up these phenomena as well. By reducing the meaning of human emotions to what is observable (based on the assumption that emotional expression is bound to emotional experience) and by quantifying these features, more encompassing features – especially those which are much harder to grasp or unquantifiable (or do not appear within given theory) – fall out of the picture. For instance, only considering the six so-called basic emotions is, to put it mildly, an oversimplification of the rich affective life of humans, as it is not considering the (social) context, the pre-reflexive dimension and the personal concerns of a concrete individual among other non-quantifiable but essential features.[13]

Not only is the emotion recognition implemented in a robot following such a reductionist approach, also the simulation of emotions follows this "rationalistic tradition". In order to interact, to have a conversation or to console it is not sufficient to recognize emotions of the other, but ultimately, we have to react adequately to them. This is the whole aim of recognizing emotions in the first place, that the robot is able to respond to the person in an acceptable manner. Thus, the robot might need to simulate feelings as well by responding with a specific tone of voice or facial expression accompanied by the adequate content of what is uttered. To illustrate this, take the following example:

> It's morning and Nicola, a 73 y.o. man, is at home alone. He feels lonely and sad since it's a long time since he last saw his grandchildren. Nicola is sitting on the bench in his living room, that is equipped with sensors, effectors and the NICA robot. After a while Nicola starts whispering and says: "Oh My …oh poor me…" (De Carolis et al. 2017: 5085-5086)[14]

13 See Eickers 2019 and Eickers & Prinz 2020 for a detailed critique of basic emotion theory and Lisa Feldman Barrett's work on the constructivist approach to emotions, recently suggesting again empirical support that the basic emotions account is not plausible (Hoemann et al. 2020).

14 NICA = "Natural Interaction with a Caring Agent"is a project which "developed the behavioral architecture of a social robot, embodied in the NAO robot by Aldebaran […]." (De Carolis et al. 2017: 5074)

What we would expect from a counterpart being present in this situation would be to understand the situation and the feelings of Nicola and to adequately engage with him. That could either mean to be compassionate, to ask and listen to him, to offer help, but also to be on edge and thus to ask him to stop complaining all day and to get ahold of himself for instance. In the given example, the researchers aim at implementing such an adequate reaction in the robot by means of simulating empathy. This is realized in the following way:

> In this scenario the voice classifier recognizes a negative valence with a low arousal from the prosody of the spoken utterance. Since the facial expression classifier cannot detect Nicola's face and expression, due to his posture, this information will not be available to NICA's emotion monitoring functionality. The evidences about the voice valence and arousal are then propagated in the DBN model and the belief about the user being in a negative affective state takes a high probability (0.74) […S]ince the robot's goal of keeping the user in a state of well-being behavior is threatened, the DBNs modeling the robot's affective mind are executed to trigger the robot's affective state. In this case, the robot is feeling *sorry-for* […A]ccording to the social emotion felt by the robot (sorry-for), the goal to pursue in this situation is console. Then, the corresponding plan is selected and the execution of its actions begins. (De Carolis et al. 2017: 5085-5086)

The authors of the paper seem to assume affective states of the robot itself by talking about "the robot's *affective* mind", that the robot "*feels* sorry-for" and that empathy as "the social emotion felt by the robot". This is striking – either not carefully written, a misleading use of metaphorical language or highly naïve, not to say just plainly wrong convictions. Endowing a machine en passant with a "mind" and affective states is scientifically untenable. But this is not the focus of my present argument. As said, I take for granted that all robots do is *simulate* affectivity. What happens in the case presented by de Carolis et al. (2017) is neither conversation nor consoling but a machine giving certain auditory output as a response to given input. In the robot's "head" and "body" nothing happens but pattern recognition. As already John Searle (1980) argued many decades ago: there is no intentionality, no semantics but only syntax for machines. The symbols and the in- and output do not *mean* anything for and to the robot. There is no difference for it whether it would "say" "I love you" or "I hate you", despite the input this is a reaction to.

The concept of care underlying the implementation just described differs significantly from the phenomenon of "caring about someone" that I have described before. In the scenario depicted here, there is no caring about a person, but a sequence of machinery output to a given input. "To care" is reduced to observable, quantifiable features and anything within the 'caring one' is held to be irrelevant. In the face of this analysis, the great public, academic, and political interest in and enthusiasm about developing so-called social or care robots at least becomes questionable, i.e. questions like the following suggest themselves: How and why does the idea gain plausibility, that a *mere simulation of the observable features of a complex phenomenon* like caring about a person is the same as actually caring about a person or at least sufficiently leading to desired results such as reduced loneliness or increased well-being? How and why does the idea gain plausibility, that anything psychological, anything within the "black box" does not matter for care to be actualized? How and why does the idea gain plausibility, that a mere "input-output" process is equal to or can substitute intentional meaningful and affective comprehension? Instead of answering these questions, I aim to illuminate their relevance for normatively assessing the implementation of care robots for the sake of reducing loneliness by considering why actual care matters, i.e. why it makes a difference if a person is truly cared about instead of (being made) believing that a simulation of care suffices.

5 Actual care matters

To care about about a person (who suffers from loneliness) by means of a robot (designed to be a companion, like Alice) is aimed at by a simulation of supposedly adequate (emotional) reactions of the robot to the (emotional) expressions of the person – by alleged interaction and conversation. The alleged interaction or conversation between a robot and a person is a mere simulation of a practice without the essential features of it being realized or even understood by the robot. In this way care or social robots can be considered a "Cargo-Cult".

> During the war they [inhabitants of the Samoan islands; IvM] saw airplanes land with lots of good materials, and they want the same thing to happen now. So they've arranged to imitate things like runways, to put fires along the sides of the runways, to make a wooden hut for a man to sit in, with two

> wooden pieces on his head like headphones and bars of bamboo sticking out like antennas—he's the controller—and they wait for the airplanes to land. They're doing everything right. The form is perfect. It looks exactly the way it looked before. But it doesn't work. No airplanes land. So I call these things cargo cult science, because they follow all the apparent precepts and forms of scientific investigation, but they're missing something essential, because the planes don't land. (Feynman 1974: 11)[15]

In the same manner, care or social robots miss 'something essential'. Independently from the emotional reactions of a person towards a robot, human-machine relationships are "sorely lacking in *some features* of human relationships" (2015: 120; emphasis added), as Troy Jollimore argues. Even if (professional) care relationships differ in crucial regards from the loving relationships Jollimore is concerned with, the *structural point* is the same: Like how the inhabitants of the island just did anything the Americans did in order to get cargo by imitating any single gesture etc., a robot can simulate all possible sayings or facial expressions without anything of their meaning being *actualized*. Caring is not realized by simulating it. One might argue that this is not problematic as such, for instance by pointing out that people might exist who claim to be fine with cargo-cult-care. Or by considering that also in human-human relationships emotions and intentions are often only pretended rather than actually experienced or meant. Although I doubt that someone really prefers a simulation over a potential actual relationship of care, this empirical question is not the point here. What I rather aim to highlight is that a simulation is just not the same as actual care. If 'something essential', as said above, is missing, it is something different. That means, even if a person might really want to (or think they want to or purport to want to) live in illusions about

15 This example is not about the specific case of the inhabitants of the Samoan islands and should not carry any colonialist undertone here. The crucial insight by making up this concept and transferring it to other realms is demonstrated by Gunther Dueck for instance, who identifies cargo-cults in science, politics and management (Dueck 2016). I use it here as a concept that helps to denote the phenomenon of simulating practices in order to reach a goal without the necessary components of the practice being understood and actualized. Another example is the hype around what in Denmark is called hygge. It is a specific way of life, an affective phenomenon that others try to reach by simulating anything Danish people supposedly do in order to have it, like having candles or cinnamon rolls around – without changing the very affective state that this phenomenon is about. For sure there is no hygge just by having some cozy cushions close by.

having a relationship in the sense of "caring about", they cannot claim, based on my analysis, to *actually be* in such a relationship of care.

This does not answer the question why actual care matters. This normative point entails (at least) two dimensions I aim to sketch here. The first one concerns the individual (i.e. their right to be acknowledged as a person rather than an object) and the second one the structural perspective (i.e. the potential societal effects of normalizing simulated care). Concerning the individual level, one could argue based on the observation that professional garegivers withdraw real emotional investments from an engagement with their patients, that it suffices also for robots to "behave *as if* they care, i.e., give the illusion that they respond to the feelings and the suffering of their care recipients" (Wachsmuth 2018). I want to counter that argument firstly by noting that the non-caring of *some* humans does not legitimize the non-caring of machines. Secondly, and this is the crucial point, such a position already takes for granted an "outcome-oriented", that is: functionalist-individualist approach that neglects, on the individual level, the import of the whole complex phenomenon of care. Even if it might be true that some humans sometimes are happy as a result of a supposed interaction with a machine (that it supposedly enhances affective wellbeing, leads to more autonomy and the like), this should not be the (only or most important) indicator for an ethical assessment of care robots as social companions. Because as others have already pointed out: "Most people do not just want to have positive feelings; they want to have 'the real thing,' which in this context is *social interaction*. Apart from its contribution to well-being, social interaction is valued as an end in itself." (Misselhorn, Pompe & Stapleton 2013: 128) And: "We don't just want to feel that we are relating to others, we actually want to *be in relationships*" (Jollimore 2015: 140; emphasis added). In this regards, it is plausible to assume that humans will not only stay lonely when someone places a robot in their home, but that this allows them realizing even more that in fact no-*body* cares about them (which is the reason for being lonely in the first place). Thus, supposedly being cared about by a robot might serve as the ultimate proof for really being lonely. This might result in learned helplessness and a shift in baselines for what is deemed to be proper care in the first place, entailing to be pleased with "mere entertainment" or just being calmed down, not even expecting to matter to someone anymore.

The structural danger that can be pointed out here is that by giving up essential features of concepts, they change their meaning because of their different use in the 'language community' (Wittgenstein). The worry is that

the meaning of what it is to engage in a caring relationship changes once we get rid of the necessary component of a counterpart being able to actually respond with their own voice. If people have 1000 or more friends on Facebook or other social media, then "friendship" cannot mean something special, intimate, something where people share important values, concerns, goals, and the like, are able to communicate in specific ways and feel specific intense feelings for one another, care about the other and assign them a high value in their own life.[16] This case of a different meaning of the concept "friend" might not be that problematic because most still seem to know the difference between their real and their Facebook-friends. But in the case of the concept of care and the corresponding practice, the danger is real, once science, research, industry and politics call for a "care of tomorrow" with which it is meant that Alice and other artifacts are really supposed to care about humans.

My normative argument for why actual care matters relies on the assumption, spelled out in section 3, that humans *should* be brought into existence as persons rather than objects and that this requires an actual counterpart being able to do this. Neither in a supposed relation between an artificial system and a human, nor between a non-caring human and another person this is actualized. Artificial systems cannot bring a person into existence as a person, they just do "not give a damn" about the other (to be clear: they cannot even not give a damn, for this intentional vocabulary just does not apply to objects). Yet, at the core of what it means to care about another person lies the conviction that humans should not be objectified but rather be acknowledged in their value and dignity. That there are many cases in which this actually is not lived, even among humans, does not mean that we should not aim for this. Implementing systems which *in principle* can not realize this, can not be the solution. A robot is not and cannot be a caring being – it is a thing, an object. But if we – we, as the society being responsible for the narratives and imaginaries which make certain solutions to societal problems visible while obscuring others – do not see the robot as an object, but rather endorse the illusion that a robot cares about us, we humanize it and objectify ourselves: We objectify ourselves by humanizing an object.[17] We allow persons to be treated

16 See also Troy Jollimore (2015) on "The importance of whom we care about".

17 It is important to remember my general point here: I do not claim that people cannot and should not have emotions for objects, as in the case of a child who has strong feelings for their teddy bear. My point is that the child cannot entertain a reciprocal relationship of care with the teddy bear. The teddy bear might be of great importance

like any other object among objects by a program which identifies the person as "human" instead of "window" or "table" by recognizing certain quantifiable patterns – and we call this "care" and "companionship". Reducing the complex process of caring into single sequences of input and output, does harm to this phenomenon and carries the risk to change it ultimately. The normalization of implementing care robots to fight loneliness bears the danger of a structural and even institutionalized (acceptance of) objectification and that we may not even expect essential features of care from humans anymore, if we get more and more used to encounters where our opposite does not care about us. This is to be phrased as a hypothetical state of affairs, as a *potential* outcome of the implementation and acceptance – the normalization – of care robots in societies. What might also get out of view in such a framing is seriously considering other possible ways to handle loneliness. It is a question of societal priorities and demands – and the narratives and imaginaries making these intelligible in the first place – whether resources are spent for the development of technologies like care robots or for structurally making professional and private human care-giving possible, acknowledged and payed properly.[18] In the conclusion I will shortly suggest to use technology for problems it is

to them and they might project intentions and feelings onto it and thus believe they entertain a relationship. I do not see any trouble here, as long as nobody makes claims for the teddy bear being able to be the care-giver for the child. In the same spirit, an elderly person might find it pleasurable to play with or cuddle the robot seal Paro. This does not appear to be problematic as long as nobody claims that Paro could or should care for and about the person – for it just cannot do this . Whether it is ethically justifiable or problematic to use robot pets for treating patients with dementia is a question not to be addressed in this paper. See Schuster (2021) for a detailed analysis.

18 This is a concrete societal, structural and political issue that goes beyond the scope of this paper, which solely aims at arguing against the possibility of robots to care and thereby wants to establish the ground for another argument against emotional bonds with robots. Here, one could argue that the implementation of care robots is not an unavoidable step because of the so-called shortage of nurses but rather a question of supply and demand and of societal priorities on how to spend time and money. Intertwined with this might be a feminist critique pointing towards the circumstance that in patriarchal societies most of the care work is still done by women. One might argue that if robots cannot care and thus, should not be considered a solution, this increases the burden on mostly women whom at worst then take care of their own children and their parents. Again, I think this worry is important but not one that emerges out of a critique of robots but rather of how societies are structured as such. That problem is completely unconnected to the idea to implement robots as care-givers.

suited for and point out the implications of the main analytical insight of my text, namely that robots just "don't give a damn".

6 Conclusion and outlook: Category mistakes and proper solutions

There have been some voices in the history of the philosophy of science drawing a completely negative picture of *technology as such*. Most prominently Karl Jaspers, Jacques Ellul, and Martin Heidegger are seen as proponents of a view that technology alienates the human and is the source of evil of different kinds (cf. also Coeckelbergh 2020: chapter 2.2). While it seems obvious that it is in many regards not a promising idea to disregard technologies as such, the general spirit of these approaches and their initial questions are of utmost importance when it comes to the normative question of robots as one specific technology, and their role in our everyday lives. How do they shape what we are and conceive ourselves to be? As robots might transform not only practices and parts of our lives but lead to "new kinds of subjects" (Foucault 1977 [1975]), we should ask the normative question of what we consider the good life to be and whether robots as care-givers, educators or lovers are conducive or detrimental to it, *before* we set these things into existence.

I argued in this paper that robots cannot care. I take this to be an important consideration that needs to be recognized in broader and general normative assessments of (emotional) bonds with robots. The main counter argument against the thesis I present here is: still robots are better than nothing and although they might not be able to care in the demanding sense I have sketched, they might enhance the well-being of the elderly – and lonely person in general – by calming them down, entertaining them etc. I do not deny that what robots, and technological aid systems more broadly, can do in the care sector is encompassing and in many regards a great achievement. Lifting people into their bed, reminding staff of the medicine their patients should take etc. are of great use. At best they not only solve the problems they are supposed to solve but also endow the human care-givers with more time and resources in order to *really* care about their patients. The problem occurs once robots are used in a realm for which they are not suited, as has been pointed out in this chapter.

Robots do not engage in conversations, and they can neither be treated as slaves nor entertain relationships. These are category mistakes. To engage

in a conversation, to be enslaved or entertaining a relationship presuppose capabilities robots do not have. We would never come to the idea that our hammer, toaster or smartphone entertain relationships with us, that we enslave them or engage in conversations with them.[19] Why do we do this with robots? Only because *we,* as their designers, intentionally made them up in a way that they *resemble* humans and that they give an allegedly similar output to that of human beings. When we compare or even equal humans with machines because there seems to be the same output to a given input, say in a conversation, we reduce these phenomena in the same way we reduce what it means to be human. We replace complex phenomena with a mere manipulation of an output that is associated with the very essence of the phenomenon. It would be much better to use specific technologies for the tasks to which they are well-suited or even more well-suited than humans in order to endow humans with the necessary time for the tasks in which their uniqueness is at issue. And this is anything that has to do with relationships, care, concern, friendship, and love. These areas are inherently human areas, because they are essentially meaningful endeavors requiring beings which are able to consciously and autonomously engage in affecting and being affected by one's counterpart. To beings which can bring their counterpart into existence as a person rather than treating them as one object among others.

My approach does not entail a prescriptive request for concrete political laws and prohibitions. It rather is an invitation to consider possible counter-imaginaries and narratives about a techno-solutionist narrative with respect to the problem of loneliness and care by arguing that humans deserve to be taken as persons rather than objects. My goal is empowerment of anybody affected by this and my intention is emancipatory and not paternalistic. Adopting a power-sensitive approach (in the sense of Foucault) to the ethics of care robots, in this spirit, means to acknowledge that society at large is affected by normalizing simulated care. This needs to be spelled out in more detail somewhere else. The invitation of the present paper is to seriously consider

19 The issue of objectophilia cannot be addressed here. What is important for my argument is the assumption that people who actually think they entertain relationships with the Eiffel tower, their smartphone or the Berlin wall (thanks to Janina Loh for making me aware of these cases) have a different concept of relationship than I am concerned with in this chapter. Neither the Eiffel tower, nor the Berlin wall, a hammer, toaster or smartphone meet the conditions to fulfill what I have spelled out to be necessary for caring and thus for such a relationship.

the ethical task to reframe the reflective capacities of individuals as users and as producers or political decision makers when considering care robots. From a perspective that does not take for granted that care robots provide a suitable means to fight loneliness, the motto of the "Alice cares" project, namely "The care of tomorrow" sounds rather dystopian. My aim was to provide grounds for hope that humans can find the courage to escape the robotization of themselves – that they enable themselves (again) to enact meaningful spaces together, to be affectable by and responsible for one another.

References

"Alice Cares Promo", March 23, 2021; https://www.alicecares.nl/media-en

Coeckelbergh, Mark (2020): Introduction to Philosophy of Technology, New York: Oxford University Press.

Coeckelbergh, Mark (2022): The Political Philosophy of AI: An Introduction, Cambridge: Polity Press.

De Carolis, Berardina/Ferilli, Stefano/Palestra, Giuseppe (2017): "Simulating Empathic Behavior in a Social Assistive Robot." In: Multimedia Tools and Applications 76, pp. 5073-5094.

Dueck, Gunther (2016): Cargo-Kulte. re:publica 2016, retrieved: May 28, 2022; https://www.youtube.com/watch?v=6YhugALYhhQ.

Eickers, Gen (2019): Scripted Alignment: A Theory of Social Interaction, Berlin: Freie Universität Berlin.

Eickers, Gen/Prinz, Jesse (2020): "Emotion recognition as a social skill." In: Ellen Fridland/Carlotta Pavese (eds.), The Routledge Handbook of Philosophy of Skill and Expertise, New York: Routledge, pp. 347-361.

Feynman,Richard (1974): "Cargo Cult Science." In: Engineering and Science 37/7, pp. 10-13.

Foucault, Michel (1977 [1975]): Überwachen und Strafen. Die Geburt des Gefängnisses, Frankfurt a.M.: Suhrkamp.

Haugeland, John (1998): Having Thought. Essays in the Metaphysics of Mind, Cambridge: Harvard University Press.

Hoemann, Katie/Zulqarnain Khan/Mallory Feldman/Catherine Nielson/Madeleine Devlin,/Jennifer Dy/Lisa F. Barrett (2020): "Context-aware Experience Sampling Reveals the Scale of Variation in Affective Experience." In: PsyArXiv, doi:10.31234/osf.io/cvjb8.

Jackson, Frank (1982): "Epiphenomenal Qualia." In: The Philosophical Quarterly 32/127, pp. 127-136.

Jollimore, Troy (2015): "The importance of whom we care about." In: Anthony Rudd/John Davenport (eds.): Love, Reason, and Will Kierkegaard After Frankfurt,

Jollimore, Troy (2015): "This Endless Space between the Words": The Limits of Love in Spike Jonze's Her." In: MidwestStudiesinPhilosophy 39/1, pp. 120-143.

London:Bloomsbury, pp. 47-72.

Misselhorn, Catrin (2021): Künstliche Intelligenz und Empathie, Frankfurt a.M.: Reclam.

Misselhorn, Catrin/Ulrike Pompe/Mog Stapleton (2013): "Ethical Considerations Regarding the Use of Social Robots in the Fourth Age" In: GeroPsych: The Journal of Gerontopsychology and Geriatric Psychiatry 26, pp. 121-133.

"Reactivity Project", March 23, 2021; https://sites.google.com/view/thereactivityproject/home.

Roethlisberger, F. J./Dickson, W. J. (1939): Management and the worker, Cambridge: Harvard University Press.

Rosa, Hartmut (2016): Resonanz, Frankfurt a.M.: Suhrkamp.

Rouse, Joseph (2002): How Scientific Practices Matter. Reclaiming Philosophical Naturalism, Chicago: The University of Chicago Press.

Sanders, K./Kianty, A. (2006): Organisationstheorien, Wiesbaden: VS Verlag.

Schuetze, Paul/von Maur, Imke (2021): "Uncovering today's rationalistic attunement" In: Phenomenology and the Cognitive Sciences 21/2, pp. 1-22.

Schuster, Kathrin (2021): Therapieroboter in der Betreuung demenzbetroffener Personen, Osnabrück: V & R Unipress.

Searle, John (1980): "Minds, brains, and programs" In: The Behavioral and Brain Sciences 3, pp. 417-457.

Sharkey, Noel/Sharkey, Amanda (2010): "Granny and the robots: Ethical issues in robot care for the elderly." In: Ethics and Information Technology 14, pp. 27-40.

Sparrow, Robert/Sparrow, Linda (2006): "In the hands of machines? The future of aged care. Mind and Machine", In: Minds and Machines 16, pp. 141-161.

Stiegler, Bernard (2010): Taking Care of Youth and the Generations, Standford: Stanford University Press.

Todorov, Tzvetan (1993 [1991]): Angesichts des Äußersten, München: Wilhelm Fink.

Tronto, Joan (1993): Moral boundaries: A political argument for an ethic of care, New York: Routledge.

Tronto, Joan (2010): Creating caring institutions: Politics, plurality, and purpose. In: "Ethics and Social Welfare" 4/2, pp. 158-171.

van Wynsberghe, Aimee (2013): "Designing Robots for Care: Care Centered Value-Sensitive Design." In: Sci Eng Ethics 19, pp. 407-433.

von Maur, Imke (2018). Die Epistemische Relevanz des Fühlens; https://osnadocs.ub.uni-osnabrueck.de/bitstream/urn:nbn:de:gbv:700-20180807502/5/thesis_von_maur.pdf.

von Maur, Imke (2021). "Taking situatedness seriously: Embedding affective intentionality in forms of living." In: Frontiers in Psychology 12, 599939.