

# TECHNOLOGY AND THE APPEARANCE OF THE GOOD: CAREBOTS, VIRTUAL VIRTUE, AND THE BEST POSSIBLE LIFE

Mark Coeckelbergh  
University of Twente

Abstract (1565 words)

Growth of the elderly population and nursing shortage place increased pressure on our health care systems. One possible response is to let care robots or carebots take over care tasks. Some of these robots appear human in some way (humanoid robots), or look and act like a pet (pet robots). As personal robots they 'share physical and emotional spaces with the user' (Cerqui and Arras 2001) and play a role in daily life. They can assist ill and elderly people by monitoring them, by delivering drugs, by moving them around, by helping them with domestic tasks. They can be used for therapeutic aims, or to entertain and accompany people.

How can we evaluate such a near-future scenario in terms of its contribution to 'the good life', given that carebots would often replace real humans or pets?

The question concerning the relation between carebots and the good life must be asked at least two levels. First, we need to address the problem of how (much) the good life depends on genuine human-human interaction as opposed to other human-human and to human-robot interactions. What is 'genuine' human-human interaction, and what is an 'appropriate' interaction between humans and technology? If we want to define 'genuine' and 'appropriate' in relation to the good life, we have to clarify what we mean by the good life. Second, we should ask what the good life means given the specific situation of ill and elderly people. Human flourishing, the good life, seems easier when one is young, healthy, and in touch with one's nearest and dearest. But these conditions are not always fulfilled. Ill and elderly people cannot aspire to the same level of human flourishing as others, and have specific needs and vulnerabilities. It is more appropriate, then, to speak about the best possible life in their situation.

On reflection, however, this term seems appropriate with regard to all of us. How absolute can the criteria for human flourishing be? Are they relative to one's personal situation and one's unique individual character, or are there needs and aspirations shared by all humans? I respond to this familiar problem in ethical theory by distinguishing between three equally important 'good life' questions that must be raised to evaluate carebots and other technology: one concerns general criteria that apply to all humans (criteria for the good life), one concerns universal criteria that apply to the particular situation of the person, the context, and the technology (criteria for the best possible life given that situation), and one concerns criteria that are unique to the person and that person's interaction with the technology (criteria for the life of person X).

In this paper, I cannot address these questions comprehensively, but I will illustrate my approach by making two arguments concerning 'good life' evaluation of human-human and human-technology relationships: one about Aristotelian friendship and one about care. Moreover, I propose that we focus on the ethical quality of the interaction between humans and between humans and carebots (and other technology), and argue that with respect to such an evaluation the robot's appearance matters rather than reality (whatever that means).

1. What kind of interaction between humans is constitutive of the good life, makes us flourish as humans? In contrast to human rights theory or the capabilities approach, the 'good life' approach does not ask about minimum criteria for human life, but asks about how to maximise the quality of that life. Part of Aristotle's answer to that question is his account of friendship given in the *Nicomachean Ethics*. For him, the most worthy relationship between people is what he calls 'perfect friendship', which requires that one desires the other for the sake of the other. It is a relationship between moral equals; we acknowledge virtue in the other. In imperfect friendship, by contrast, the other is a source of advantage or pleasure; it is a morally asymmetrical relationship.

This notion of friendship, however, is far too demanding to be a model that could guide the interaction between care professionals and care receivers. At the same time, we do not want to resort to the 'minimum' morality embodied in ethical codes, contract theory, and care protocols; we need a concept that refers to 'warmer' and less instrumental relationships, and one that concerns the good rather than (only) the right. A further problem is that Aristotle's theory of friendship has nothing to say on human-technology relationships; it only deals with humans and it seems inappropriate to talk of moral (a)symmetry between humans and robots, or to discuss a robot's moral development. Moreover, Aristotle's theory appears to be addressed to active, rational, independent, and healthy male citizens of high social rank, whereas ill and elderly people

often embody dependency, vulnerability, inactivity, and non-participation in the public life. Perhaps they even lost their rational capacity.

Some of these problems could be avoided by appealing not to Aristotelian friendship but to companionship, which has something of the 'warm' connotation of friendship, but is less demanding. It does not require symmetry, without supposing a totally instrumental relationship. This makes the term suitable to apply it to interactions between humans and non-humans such as animals, pet robots, and humanoid robots. People sharing their physical and emotional space with pets or personal robots do not typically experience their interaction with the (robot) pet or (the future) humanoid carebot in instrumental terms. Rather, empirical research suggests that these non-humans appear as their companions, and are interacted with accordingly. At the same time, the humans are aware of the 'real' asymmetry. The contribution such interactions make to human flourishing, therefore, must be judged on the basis of what the people experience. Appearance turns out to be ethically relevant, not 'reality', whatever that means.

But if appearance matters so much for the good life, can we push this argument further, and consider the possibility of symmetry? An advanced humanoid robot could appear to us as a morally equal companion, in spite of our knowledge that it is a robot. In that case, we would ascribe 'virtue' to the robot on the basis of appearance. A robot with virtual virtue we would call a 'good' robot. This is less odd than it may seem at first sight. After all, we are used to virtual characters in novels, films, and computer games. In fact, very little is necessary for us to ascribe a personality to an object such as a computer (Reeves and Nass 1996).

2. We must develop criteria specific to the kind of life ill or elderly people lead, who do not meet the Aristotelian assumptions regarding autonomy and independence, and can only attain the best possible life given their vulnerable situation. Such an ethics of the best possible life could put forward the requirement that ill and elderly people enjoy interaction with others who care for them or who appeal to their capacity to care. Here 'care' does not have the narrow meaning of 'medical care' or 'health care', and it is not required that one desire the other (as in Aristotle's perfect friendship); rather, one desires the best for the other in the given situation of illness or old age and given that person's unique character and identity. I believe this criterion of 'careship' can be met by others than (close) friends and relatives only, but is more demanding than what protocols and codes require from health professionals.

However, while this criterion could, in principle, be met by health professionals if certain contextual conditions are fulfilled (e.g. sufficient staff and time available), they cannot be met by robots. They could 'deliver' care but not 'really' care about their patients. However, if we remove the reality demand, we create an opening: carebots could appear to care. I do not suggest that people should be fooled; rather, people are usually aware that the robot is a robot, they know it, but they would interact with the robot on the basis of its apparent pet-ness or human-ness, on the basis of what they feel, and this could contribute to their human flourishing. Personal robots are never 'mere' robots in this sense. They are what Turkle, Taggart and others call 'relational artefacts' in their study of human-robot interaction in nursing homes (Taggart et al. 2005). In the study, a pet robot looking like a baby seal was treated as if it needed nurturing. This suggests that if people feel that the robot needs their care or that the robot cares – with care understood more broadly than 'medical care' – then that matters to their human flourishing, to their best possible life given their situation and character. Both human care and robot care should be adapted to the specific situation ('case') of the person and the kind of human-robot interaction that emerges between the unique, individual person and the robot. For example, Turkle's research shows that people respond very differently to the same robot, according to their character.

I conclude that in order to evaluate carebots in terms of the good life, we need develop three kinds of criteria: general/human, universal/situation-specific, and individual-unique criteria. I have discussed friendship, companionship, and careship as models for human-human and human-robot interaction. I call for more attention to empirical research on such interactions to further develop reflection on how people in health care contexts could attain the best possible life, that is, the best possible life for humans in that situation and being that unique individual.