

## Social Robots and Digital Well-Being

**Citation for published version (APA):**

Dennis, M. (2022). Social Robots and Digital Well-Being: How to Design Future Artificial Agents. *Mind & Society*, 21(1), 37-50. <https://doi.org/10.1007/s11299-021-00281-5>

**DOI:**

[10.1007/s11299-021-00281-5](https://doi.org/10.1007/s11299-021-00281-5)

**Document status and date:**

Published: 01/06/2022

**Document Version:**

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.tue.nl/taverne](http://www.tue.nl/taverne)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[openaccess@tue.nl](mailto:openaccess@tue.nl)

providing details and we will investigate your claim.



# Social robots and digital well-being: how to design future artificial agents

Matthew J. Dennis<sup>1</sup> 

Received: 19 September 2020 / Accepted: 13 September 2021 / Published online: 24 September 2021  
© The Author(s) 2021

## Abstract

Value-sensitive design theorists propose that a range of values that should inform how future social robots are engineered. This article explores a new value: digital well-being, and proposes that the next generation of social robots should be designed to facilitate this value in those who use or come into contact with these machines. To do this, I explore how the morphology of social robots is closely connected to digital well-being. I argue that a key decision is whether social robots are designed as embodied or disembodied. After exploring the merits of both approaches, I conclude that, on balance, there are persuasive reasons why disembodied social robots may well fare better with respect to the value of digital well-being.

**Keywords** Social robotics · Digital well-being · Embodiment · Disembodiment

## 1 Introduction

Digital well-being will become an increasingly important measure of the quality of life in the 21st century. While this topic first emerged as a concern over social media use—especially screentime—increasingly theorists are employing the idea to evaluate our interactions with technology more widely (Burr et al. 2020; Floridi 2014; King 2019; Snow 2019). Understanding digital well-being involves picturing what the contemporary good life looks like, and focusing on how emerging technologies can best enhance human flourishing. Theorists who work on this topic apply character-orientated ethical theories to the technologies we live and interact with on a daily basis, showing how digital artefacts and services can contribute to living a good life, to cultivating character, and to achieving happiness or other eudaemonic states. Christopher Burr et al.’s literature review on recent research on digital well-being illustrates this approach. Burr begins by stating that digital well-being is ‘what it means to live a life that is *good for* a human being in an information society’, after

---

✉ Matthew J. Dennis  
m.j.dennis@tue.nl

<sup>1</sup> TU Eindhoven, Eindhoven, Netherlands

which he and his co-writers move to show how this concept should inform our evaluation of future technologies (Burr et al. 2020, p. 1). Scholars who are sympathetic to prioritising digital well-being seek the resources to apply it to technology from diverse ethical theories of character and virtuous character traits: from neo-Aristotelianism (Coeckelbergh 2009; Elder 2019; Vallor 2011, 2012a, 2016) to Stoicism (Dennis 2019; Klinecwiez 2019), from Confucianism to Buddhism (Vallor 2016; Wong 2016), to figures in post-Kantian philosophy (van de Poel 2012; Verbeek 2012). While each of these traditions propose their own conceptions of human flourishing,<sup>1</sup> they are united in their view that the good life includes appropriate social relationships—the most prized of which is friendship. Consequently, the capacity of social robots to provide either: (1) fully fledged virtue friendship (Danaher 2019), or (2) task-orientated companionship (De Graaf 2016; Elder 2019; Vallor 2012b)—sex robots, for example (Danaher and McArthur 2017; Devlin 2018; Peeters and Hasegag 2019)—should be considered a vital issue in social robotics.

This article evaluates the future morphology of social robots using a character-orientated framework, focusing on their ability to contribute to the good life by providing both friendship and social companionship. Instead of analysing social robots in terms of the specific moral harms they may (or may not) cause, my approach takes inspiration from Mark Coeckelbergh's observation that the 'focus on the *morality* of robots means that *ethical questions* concerning how humans interact with [them] remain out of sight' (2009, p. 217; both emphases added). To do this, I engage with the empirical literature on this topic (where available), although my primary aim is to prepare the conceptual ground for an account of how we should design social robots in ways that prioritise digital well-being. Approaching social robotics from a digital well-being perspective involves focusing on the following questions: what design choices will facilitate living well with social robots? And how will these design choices impact on our digital well-being?

In what follows, I argue that we can increase the capacity of social robots to exist harmoniously with the demands of our digital well-being by ignoring media hype and remaining open-minded about their likely future morphology. This means that we have to interrogate the design choices made by the creators of today's social robots at the most elementary level. I begin by exploring the various forms that future social robots might take, emphasising that both philosophers of technology and the public often find it hard to predict the direction that future developments in technology may move. After this, I compare the advantages of embodied social robots (ESRs) and disembodied ones (DSRs).<sup>2</sup> To do this, I focus on what the character-orientated ethical traditions agree is a core aspect of the good life: friendship,

<sup>1</sup> It is important to recognise cross-cultural differences in the values we expect technology to embody (Bruno et al. 2019, Shaw-Garlock 2009). Coeckelbergh emphasises this. He writes, 'some ways of life are better than others and that some goods are good for all humans', but he avoids the term 'the good life' because he argues that there 'are many ways of living that can be called good' (2009, p. 220). While I agree that the term might connote an overly determining sense of the good to modern ears, Aristotle and neo-Aristotelians typically emphasise that the good life comprises plural goods that differ according to the individual concerned. See Aristotle's discussion of Milo the wrestler in NE 1106b.

<sup>2</sup> I adopt an inclusive notion of DSRs, including AI-enabled chatbots and virtual assistants.

and similar types of beneficial social relations. In conclusion, I lay out four key issues that circumscribe what we should consider when thinking about how to live well with social robots.

## 2 Picturing tomorrow's artificial agents

Our inability to accurately predict the future path of emergent technologies often precipitates retrospective amusement. Take the hit 1960s cartoon, *The Jetsons*, for instance. Although a few of the technological innovations that the show depicts have come to pass (the family's automated hoovering device resembles a Roomba, for example), there is no sign that life in 2062—the year the cartoon ostensibly depicts—will turn out as the creators imagined. In fact, perhaps the most entertaining thing about watching *The Jetsons* today is witnessing how naïve the technological utopia of the sixties turned out to be, combined with its woefully unimaginative depiction of gender roles. So are we in danger of entering a similar imaginative dead-end in our collective vision of future social robots? Are the ways in which we envisage their development—either in the popular imagination or by more measured expert predictions—likely to depart from how they actually develop in future decades? Given the current ways that the media depicts social robots (which I believe filters into the philosophical literature on this topic), I believe that we should answer 'yes' to both questions. While this is likely to apply to many aspects of social robotics, I contend that it is most salient when we consider the *morphology* of future robots, specifically whether they will be designed as embodied creatures (whether humanoid or not)<sup>3</sup> or disembodied ones. The main reason for this, as I hope to show in what follows, relates to how the physical form of a social robot determines how well it can align itself with the promotion of our digital well-being.

In recent years, two differing visions of the likely form of future social robots have been offered by blockbuster films. In the most recent, *Ex Machina* (2015), writer and director Alex Garland envisions social robots as fully embodied AI-powered companions. After examination, the main character, Ava (played by Alicia Vikander) not only shows the ability to pass the Turing Test with flying colours, but eventually becomes a sought-after romantic partner for the protagonist, Caleb (played by Domhnall Gleeson). Garland envisages Ava in line with the dominant tradition in social robotics, depicting her, and her robot sisters, as effectively indistinguishable from human beings. Not only does Ava demonstrate general artificial intelligence and sophisticated emotional capacities, but her body is so functional that she can have sex and (so her creator claims) even take pleasure from it. The film

<sup>3</sup> This article aims to contribute to the question of whether ESRs or DSRs are most compatible with digital well-being, but there is the further question of whether humanoid ESRs or non-humanoid ESRs are most compatible with it. The second question may well be easier to resolve, however, and there are already several studies examining the particular merits of humanoid ESRs over non-humanoid ones. These typically examine the ability of non-humanoid ESRs to express emotion, and it is interesting—and perhaps counterintuitive—*Machina* (2015) that it is DSRs that fare better at this. See Novikova and Watts (2014) and Lakatos et al. (2015).

ends with Ava's escape from the laboratory, leading the viewer to confront the ethical questions that would arise if social robots such as Ava were widely deployed. We are invited to consider a 'near-future scenario', as Coeckelbergh describes the advent of social robotics, in which 'living with robots will be as habitual as living with TV, mobile phones, and internet' (2009, p. 217). So how would it be to live well with ESRs of Ava's sophistication? What ethical questions would living with such social robots precipitate? Ava's depiction in the film suggests that the answers to these ethical questions will be complicated, confusing, and could undoubtedly upset many of our cherished assumptions about humanity's role in a world that it starts to share with social robots.

Nevertheless, it is my contention that we may never need to answer these questions. This is for two reasons; the first contingent on how future technology develops, the second more substantive and conceptual. The first reason (1) is that current advances in robotics show little sign that ESRs of Ava's calibre—autonomous, general artificial intelligence that is integrated with sophisticated tactile and emotional abilities—will occur for many generations (if at all). Of course, the history of development of robotic technology has shown itself to be unpredictable (Moravec 1988, Ch. 1), so it is difficult to predict with accuracy how far away ESRs at Ava's level of sophistication are likely to be.<sup>4</sup> What we do know, however, is that while the various technologies that constitute the embodied aspects of Ava are improving, they are doing so at a far slower rate than non-embodied ones. The AI functionality of social robots is developing far faster than their locomotion, tactility, and human-like appearance. Over the last twenty years our ability to furnish AI with large-scale data has led to a series of astonishing developments. These advances have powered various kinds of artificial assistants (digital assistants, chat-bots, etc.), which are still going from strength to strength. By comparison, ESRs that successfully integrate these technologies are still in their infancy. For example, much industry attention has been paid to getting social robots to surmount the infamous 'uncanny valley', for instance, but this problem has not been fixed (Mori et al. 2012; Mori 1970). Until the point that humanoid robots are—like Ava—*perfectly* convincing, the problem of the uncanny valley is in danger of becoming ever more entrenched.

The second reason (2) is a conceptual one pertaining to our digital well-being. Even if an ESR like Ava was technically possible, living with such a creature may detract from our experience of the good life in ways that would not occur with a DSR. This is because, as I hope to show in the next section, embodiment introduces complications for our digital well-being, as well as making unwarranted assumptions on the likely future development of robotic technology. In other words, our digital lives with social robots may be better if these robots that are *not* embodied in the way that Garland envisages Ava to be. While the goal of future digital well-being

<sup>4</sup> Over thirty years ago at the time of writing, Hans Moravec famously stated that he believes that 'robots with human intelligence will be common within fifty years' (1988, p. 6). He notes that what is challenging for robots (tasks relating to locomotion, for instance) is simple for humans, whereas what is hard for humans (tasks relating to computation, say) is easy for robots. More recently, consider the ambitious claims made by Hanson Robotics ([www.hansonrobotics.com](http://www.hansonrobotics.com)) for their 2016 humanoid robot, Sophia, as well as the subsequent media contestation of these claims. Thanks to Madelaine Ley for these references.

is likely to be served by some kind of social robot technology, by only imagining these agents as embodied, we may be neglecting both how emerging social robots will *actually* develop (due to overestimating the likely improvement of existing technology)—and ignoring the morphology that these robots would *best* take up. Nevertheless, the powerful sketch that Garland presents us provides a useful jumping-off point, especially when we compare it with very different depiction of a social robot that has also been the subject of a recent film. This brings us to our next example.

Spike Jonze's depiction of a social robot in his 2013 film, *Her*, narrates the story of a romantic relationship between the protagonist, Theodore (played by Joaquin Phoenix), and a DSR called Samantha (ventriloquised by Scarlett Johansson). Samantha exhibits the core features of a social robot, and the power of her AI functionality causes Theodore to fall in love with her. Not only does Samantha exhibit astonishing general intelligence, intuition, creativity, and emotional sensitivity, she is plausibly depicted as a key source of well-being in Theodore's life. He *connects* with her. He shares cultural, emotional, and intimate references with her. And—on the surface at least—his life seems to *flourish* through what he comes to regard as his primary social relationship. The reason for this is that Samantha meets all the criteria that are adduced for both a well-chosen life partner and a competent social robot. In her pioneering work on social robotics, for example, Cynthia Breazeal writes that:

[A] sociable robot is able to communicate and interact with us, understand and even relate to us, in a personal way. It should be able to understand us and itself in social terms. We, in turn, should be able to understand it in the same social terms – to be able to relate to it and to empathize with it. Such a robot must be able to adapt and learn throughout its lifetime, incorporating shared experiences with other individuals into its understanding of self, of others, and of the relationships they share. (2002, p. 1)

Jonze depicts Samantha doing all of these things, invariably with excellence and finesse. Nevertheless, it is significant that the first challenge to their love occurs when Theodore and Samantha attempt to consummate their relationship. Since this requires embodiment, Samantha employs a sexual surrogate, which is when the critical question of whether a DSR can fully replace a human being as a romantic partner comes to the fore. Breazeal also notes this. She tells us that there are 'significant differences between the physical world of humans and the virtual world of computer agents', the most striking being 'physical and immediately proximate interactions' (2002, p. 5). Yet sexual relationships are only one of many kinds of social relationships that contribute to human flourishing, and these kinds of relationship are sometimes better described as instrumental as they are (often) directed towards mutual-pleasure seeking, rather more elevated types of friendship.

While embodiment may be required for sexual relationships, it is hard to defend the idea that it is a necessary part of *all* rich and meaningful social interactions, as I show below. Samantha, for example, fulfils all of the conditions that Breazeal views as required for being a social robot, without the far-flung technological leaps that an ESR such as Ava would require. On the one hand, while Ava fulfils more criteria for humanness than Samantha precisely because of her embodiment (in fact Theodore

suggests that it could be an extra requirement of the Turing Test), this comes at the cost of requiring us to postulate many radical jumps in technological development, even technologies that currently only exist in the most speculative or rudimentary forms (Bayram and Ince 2018; Meghdari and Minoos 2018; Raj et al. 2018). On the other, Samantha offers us a vision of the future of social robotics that at current rates of AI development (Bughin et al. 2017)<sup>5</sup> are likely to be achieved in the near-term future (within 5 years), rather than in decades. Leaving aside the arguments based on the likely development of technology, in the next section I move to explore the palpable advantages of a DSR, such as Samantha. Living well with artificial agents, I suggest, favours DSRs precisely because lack of embodiment creates a range of opportunities for artificial agents to add to and augment the good life, especially our digital well-being. Might not DSRs contribute to the good life insofar as they enrich our lives more widely? I think that there are persuasive reasons for this view, so in the next section I explore the extent to which many cherished human relationships are already *entirely* mediated by technology.

### 3 Social robots: to embody or disembody

Fuelled by Hollywood, our collective imagination often envisages social robots as embodied—typically humanoid—creatures. I have argued that Ava epitomises this. She is indistinguishable from a human being, and a substantial advance on earlier depictions of ESRs, such as R2D2 or C3PO. This is not to say that an ESR like Ava needs to exist in humanoid form. There is now a wealth of scholarship indicating that ‘asymmetrical sociality’ with non-humanoid robots also has strong benefits (Cerulo 2009; Hakli 2014). This might lead us to view cultural tropes like Ava as irrelevant to how future social robots will in fact develop. Nevertheless, so far I have argued that these tropes distort our predictions about how robot technology will progress, both causing us to overestimate the sophistication of near-term technologies and weakening our ability to evaluate these advances from the standpoint of digital well-being. Nevertheless, there are obvious reasons to think that a DSR might be *less* compatible with the 21st century good life than an embodied one. These reasons are important to scrutinise and ultimately—I contend—to eliminate. This is the task of the remainder of this section, after which I explicate the reasons why DSRs might be *more* compatible with our digital well-being in the final one.

If character-orientated traditions cited above are right to think that sociality is an important part of human flourishing, then it is understandable that we think of sociality analogously to how it has been traditionally understood: as meaningful interaction with other human beings. Recent research goes some way to supporting this

<sup>5</sup> In 2017, consultants at the McKinsey Global Institute made predictions for AI development in several key fields (‘retail, electric utility, manufacturing, health care, education’). Looking at how AI has improved in each of these areas in the four years since the report was published, shows that growth is in line (or has exceeded) these predictions. I contend that this also applies to the use of AI in DSRs, even if this does not proceed at the rate that the makers of *Her* suggests.

intuition. I have already noted Breazeal's emphasis on the 'significant differences between the physical world of humans and the virtual world of computer agents'. Similarly, in their empirical testing of engagement with fully embodied robots, Silvia Rossi et al. found that:

When a person interacts with an embodied physical agent, he/she is typically more engaged and influenced by the interaction with respect to other technologies, and it has been shown that subjects are more likely to value their experience as more satisfying when confronting with human-like interfaces. (Rossi et al. 2018, p. 265; see also Mataric 2017).

*Mutatis mutandis* it is easy to agree with Rossi et al. that ESRs would be 'more engaging than a 2D virtual agent' (2018, p. 266). Given what I have said about the importance of sociality—specifically virtue friendship—for human flourishing, such engagement may well equate to a greater capacity for sociality, which in turn would potentially increase our digital well-being.

Nevertheless, in Sect. 2, I argued that we should resist Rossi's conclusion for two reasons: first, because if we apply this insight to our designs of social robots, then we have to make unwarranted assumption that the technology powering future ESRs will improve significantly. This is overly ambitious, as several empirical studies suggest (Bughin et al. 2017; Bayram & Ince 2018; Meghdari & Minoo 2018; Raj et al. 2018). These studies concur that the toughness of technological challenges give us little reason to think that advances in robotics will allow embodiment in convincing humanoid form. Second, even if such advances do occur, they may come at the cost of reducing our flourishing in other ways, specifically because embodiment may turn out to present new challenges for our digital well-being that a DSR avoids. But cultural depictions of social robots can lead us in the right direction as well up the wrong track. What is significant about Samantha—the artificial agent in *Her*—is that she fulfils the key functions of a social robot *without* embodiment, only missing these when she is required to become a sexual partner. She presents a constellation of personal character traits, intellectual virtues, and an emotional profile that (at least at first) demonstrably enhances Theodore's life. So what hangs on embodiment? What should lead us to think that existing technology that is compatible with, or actively promotes, our digital well-being requires embodiment at all?

One way to approach these questions is to think of technologies that many of us already use to enhance our practical lives that do not require embodiment at all. Precedents concerning how we already use online technology to maintain friendships suggest that a DSR that actively contributes to human flourishing is conceivable in terms of existing technology. Speaking (with or without video) to loved ones in far-flung parts of the globe using an earpiece and a microphone has become a widespread activity. Similarly, many relationships that are primarily sustained over text message have been shown to be valuable and highly meaningful for both parties.

Supporting this, some scholarship has recently argued that (1) human relationships that are exclusively mediated by technology can be robust and can contribute to our digital well-being (Elder 2014, 2019), as well as (2) the extent to which human beings can form virtue friendships with ESRs (Danaher 2019). Elder argues that we overestimate the importance of embodiment when evaluating the status of



friendships we form on social media. For Elder, the friendships that we maintain on social media can be considered to be virtue friendships, the highest grade of sociality on the Aristotelian schema. Online friendships, Elder claims, allow us to manifest all the essential aspects of virtue friendship, including the ‘distinctively human activities such as conversation and exchange of thoughts, mutual development of ideas, making art, and playing games’ (2014, p. 287). Moreover, these friendships are none the worse for *not* having a physical dimension. While it might be detrimental if *all* our friendships took place online, it is permissible for *some* of our friendships to be mediated by online technology (2014, p. 293). From this Elder concludes that physicality in friendship is not required.

Elder’s work is still less radical than the ‘robot friendship thesis’ proposed by John Danaher (2019, p. 1), cited above, as Elder envisages a friendship as involving two *human beings*, albeit sustained through online technology. Danaher notes that a generation of robot ethicists have poured much scorn on the idea that fully fledged friendship is a relation that we may be able to sustain with robots (Danaher cites: Sparrow and Sparrow 2006). Danaher’s own work is an exception to this trend. He argues for two claims *à la* Aristotle: first, that the highest grade of friendship—virtue friendship—can occur, in principle at least, between a human being and a social robot, so that ‘robots can be plausibly considered our *virtue* friends’ (2019, p. 2). Second, he argues, more modestly, that ‘robots can be our *utility* or *pleasure* friends’, and that they can ‘complement and enhance more ideal friendships with other human beings’ (2019, pp. 5, 13).

Both these claims help support the idea that social robots are most compatible with our digital well-being when they are disembodied, that is, when they follow the model of Samantha instead of Ava. On the one hand, Elder suggests that embodiment is not required for the virtue ideal of friendship, the highest grade of friendship according to the Aristotelian scheme. On the other, Danaher gives us reason to think that such a friendship can occur when it takes place with an artificial agent, not just when technology mediates the relationship with another human being, in the way that Elder imagines. Taken together, both positions create the space in which to evaluate social robots that have been designed with digital well-being in mind. It makes sense to think of robots as enhancing our sociality, certainly in terms of their ability to offer companionship, but potentially in more meaningful ways such as virtue friendship too. Similarly, there is no requirement for future robot companions to be embodied fulfil this role. Many highly meaningful friendships are already conducted online, so a DSR would merely be one of many kinds of virtually mediated relationship.

In Sect. 2, I examined two strikingly different visions for future social robotics that have been offered in recent films. In this section, I argued that these cultural visions may have permeated academic discussion on this topic, circumscribing how some scholars imagine the future direction of social robotics (Rossi et al. 2018; Mataric 2017). Nevertheless, I sided with both Elder’s and Danaher’s claims that we have reason to think that friendship—perhaps even virtue friendship—neither requires embodiment, nor disqualifies the participation of artificial agents. Both insights are useful in thinking about the morphology of future social robotics. Finally, since advances in AI constantly outstrip technological developments in

ESRs, there is further reason to think that a DSR that actively contributes to our digital well-being will be feasible long before an embodied equivalent. So, in addition to the attraction of near-term technological feasibility, what are the advantages of *disembodied* social robots over *embodied* ones? Most importantly, are there other reasons to suggest that a DSR would be *more compatible with our digital well-being* than an ESR?

## 4 On the advantages of disembodied social robots

In the spirit of Coeckelbergh's claim (cited above), I have argued that the broader ethical questions of social robotics often receive less scrutiny than the more narrowly focused moral ones.<sup>6</sup> One of these is the capacity of social robots to contribute to our digital well-being by exercising our social capacities, an aspect of the good life that all character-orientated traditions agree is vitally important. The next 4 subsections sketch the advantages of DSRs in terms of their capacity to contribute to our digital well-being.

### 4.1 ESR costs: technological, financial, and ecological

The kind of reasons that mitigate against imagining the future of social robots as ESRs are negative ones insofar as they are directed towards the high costs associated with embodied robot technology. These high costs are related to the *embodied morphology* of ESRs, so they do not apply to DSRs by definition. I have shown that social robots have much potential to benefit our digital well-being. This means that we have reason to seriously pursue the project of social robots, so the reasons *against* ESRs that I address in this section count as reasons *for* DSRs, if these reasons relate to problems explicitly concerning embodiment. Since an ESR requires capabilities relating to embodiment (locomotion, tactility, gait, etc.) *in addition* to the technological specifications it shares with DSRs, then creating an ESR is far harder technological challenge (1) and comes with higher (2) financial and (3) ecological costs. These present ESRs with a set of pressing problems. I sketch each of these challenges in order below.

We encountered the radical mismatch between technological progress and cultural expectations in Sect. 2. Technologies relating to embodiment, compared to those that simulate mental processing power, are progressing markedly slowly. I have noted that Moravec's early prediction still holds: robot locomotion is an especially tough challenge (1988, pp. 25–29). I have also noted—*pace* Hanson Robotics' Sophia—that today's ESRs have been unable to scale the uncanny valley (as Ava arguably does). Overcoming these challenges would involve much expenditure of

<sup>6</sup> For philosophers who use this distinction in a wide and a narrow sense, see John Skorupski (1998), Charles Taylor (1991: 4–5), Bernard Williams (1993 [1972]: xiv, 73–81, 1985: 6–8, and 2008 [1993]: 41–2).

technological expertise, and it may well be that some of these challenges cannot be surmounted without paradigm shifts in current technological processes.

Closely associated with these technological challenges are those pertaining to the *financial* costs of an ESR. Developing robotic technology is expensive in general, but the costs associated with embodiment are disproportionately large. While we have seen that ESRs capture our imagination on the big screen, unless they benefit our digital well-being in tangible ways, then these costs may well be prohibitive. Given how ESRs present us with many pressing technological challenges, from a financial point of view we would do better to divert our resources towards the development of DSRs, if these prove to be more cost effective.

Finally, we must not think costs in purely financial terms. The drain on environmental resources of an ESR—to initially produce it, to maintain it, and ultimately to superannuate it—means that it is also highly costly in *ecological* terms. An ESR like Ava would need to be regularly maintained precisely because she is embodied. Her body parts will suffer wear and tear, her battery will need replacing, etc. While her non-embodied intellectual capacities might routinely need upgrading, this would have next to no ecological effect compared to upgrading the embodied parts of her.

## 4.2 Streamlining digital technology

Recent studies of digital well-being, such as Burr et al. focus on the disadvantages of an overabundance of technology in our practical lives (2020, p. 16). While ESRs bring potential benefits to our digital well-being, these benefits come at the cost of saturating our environment with yet more technology. In the case of an ESR, this technology is not only costly in technological, financial, and ecological terms, but it adds another layer of technological mediation to the contemporary world.

By contrast, streamlining technological intrusion in our immediate environment has the potential to improve our digital well-being. Precisely because she is more humanlike, interacting with an ESR of Ava's calibre may well not constitute the same threat to our digital well-being as excessive screentime, for example, but doing so may lead to new problems. It is easy to imagine a scenario in which early incarnations of ESRs are less than perfect. Of course, before *fully* functional ESRs are developed, this scenario is likely (even the fictitious Ava has many previous prototypes). In this case, it is easy to imagine the kinds of tortuous interactions we might be made to suffer at the hands of an ESR that was not working perfectly.<sup>7</sup> While Hanson Robotics is to be congratulated on pushing ESR technology to new heights, Sophia's public appearances typically show that it is eventually burdensome to interact with Sophia because her functionality is more rudimentary than either Ava's or a human being's. In terms of her disembodied functions, Sophia quickly reaches the

<sup>7</sup> This will apply to DSRs too, as those who have tried to unsuccessfully communicate with a virtual assistant will have experienced first-hand. Nevertheless, as the technological challenges of creating fully functioning ESRs are greater than those that apply to DSRs, then we have reason to think that DSRs will be fully functioning significantly sooner than ESRs. Thanks to an anonymous reviewer for pressing me to expand on this point.

limits her comprehension or her ability to meaningfully contribute to a conversation; in terms of her disembodied ones, it is not clear that she has managed to escape the uncanny valley.

### 4.3 Ubiquitous friendship and companionship

In Sects. 1 and 3, I argue that friendship is rightly celebrated as a vital dimension of the good life by character-orientated ethical systems from culturally diverse origins. Social robots have the potential to radically increase the availability and access we have towards friendship and companionship. Furthermore, since they are technological artefacts, social robots are well understood as contributing to our digital well-being in particular (as opposed to human flourishing in general). Extrapolating the findings of Danaher (2019) and Elder (2014, 2019) illustrates this. On the one hand, Danaher argues that establishing social relationships with robots—either companionship or fully fledged virtue friendship—is possible, and that it contributes to our capacity to lead a flourishing life. On the other, Elder argues that the relationships that we maintain on social media meet the criteria for friendship, despite these relationships not involving embodied, physical presence. Taken together, these findings show that relationships with DSRs could count as friendships—or at least rich companionships—and therefore could contribute to our capacity to lead a flourishing life. Taking this as our starting point clarifies the advantages of DSCs. This technology massively increases the ease with which we form empowering social relationships, which in turn contributes towards our flourishing. In other words, DSRs have the potential to increase the *ubiquity* of social robot friendship.

I have already introduced two reasons for this: ESRs are less expensive (4.1) and they are easy to incorporate into a streamlined digital life (4.2). Finally, there is another reason: our relationship with DSRs can potentially be maintained in a much wider range of situations, increasing the times that we can benefit from their social input. Lack of embodiment means that we can develop our relationship with an DSR while swimming, riding a motorcycle, or when we seek solitude but also want to have access to a confidant. It is easy to imagine how a DSR could be integrated into an existing smart device or into an earpiece. Many of us already use these kinds of technology to mediate our friendships and to access other kinds of informational content; a DSR could function in precisely the same way. This means that, although a DSR is not present at any particular location, it is effectively present *everywhere*. Samantha illustrates this well. Theodore is able to communicate with Samantha in a wide range of situations (including while swimming). He has a constant and ubiquitous companion, which allows him to reach a high level of emotional and intellectual intimacy. What Samantha loses embodiment, she makes up for in ubiquity.

### 4.4 Ethical advantages

While in an earlier paper Danaher notes that even the use of disembodied AI assistants can be ‘dehumanising’, lead to ‘cognitive degeneration’, and can even ‘degrade important interpersonal virtues’ (2018, p. 629), the ethical issues attending ESRs

are arguably more grave. Ethicists have already raised important concerns about the exploitation of ESRs, especially in the context of sex robots (Devlin 2018; Peeters and Haselag 2019). One of these issues is consent. Precisely because of their embodiment, ESRs can be engaged with sexually, even in contexts where a human being may have reason to resist (humiliating or degrading contexts, for example). Danaher (2018) is right to say that we should have ethical concerns about the mistreatment of DSRs, but these concerns quickly multiply once embodiment is involved. As Janna van Grunsven and Amiee van Wynsberghe note, the embodiment of sex robots both allows them to simulate emotions and also makes them vulnerable to abuse (online first; 2019).

Nevertheless, we would do well to heed Danaher's warning about the dangers of thinking DSRs such as chatbots are immune to the risk of abuse. Not only does this apply to our interactions with these artificial agents, but it applies to the very constitution of the agents themselves. Take gender, for example. The gender role that a DSR may express presents us with a range of serious ethical issues. Take, for instance, the ways in which DSRs respond to commands is important. If a DSR is designed to encourage prosocial, even virtuous behaviour, then we need to ensure that it has the capacities to actively discourage rude or offensive behaviour. Theodore is always courteous when addressing Samantha, but we need to consider a possible world in which he was not as well disposed.

## 5 Conclusion: social robotics and digital well-being

This article has argued that social robots have much potential to benefit our digital well-being, especially if we resist imagining their future morphology in standard terms. Reflecting on how to design social robots that prioritise our digital well-being indicates that we should carefully consider the advantages of DSRs, rather than following the current cultural (and often academic) trend that primarily conceives them as embodied. Moreover, designing with digital well-being in mind has a further advantage: although DSRs are less ambitious than ESRs (and perhaps capture our imagination less readily), so far they have developed at a swifter pace than their embodied brethren. This means that what is technologically feasible may be *more* compatible with our digital well-being than what is (as yet) technologically difficult. Unlike areas of technological innovation, the practical constraints on the development of ESRs bodes well for our ability to live well with DSRs in the years ahead.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Bayram B, İnce G (2018) *Advances in Robotics in the era of industry 4.0: managing the digital transformation*. Springer, Cham
- Breazeal C (2002) *Designing sociable robots*. The MIT Press, Cambridge, Massachusetts
- Bruno B, Recchiuto C, Papadopoulos I et al (2019) Knowledge representation for culturally competent personal robots: requirements, design principles, implementation, and assessment. *Int J Soc Robot* 11:515–538
- Bughin J, Hazan E, Ramaswamy S, Chui M, Allas T, Dahlström P, Henke N, Trench M (2017) *Artificial intelligence: the next digital frontier?* McKinsey Global Institute, New York
- Burr C, Taddeo M, Floridi L (2020) The ethics of digital well-being: a thematic review. *Science and Engineering Ethics* (Online first: unassigned to a volume or issue)
- Cerulo K (2009) Non-humans in social interaction. *Ann Rev Sociol* 35(1):531–552
- Coeckelbergh M (2009) Personal robots, appearance, and human good: a methodological reflection on roboethics. *Int J Soc Robot* 1:217–221
- Danaher J (2018) Toward an ethics of AI assistants: an initial framework. *Philos Technol* 31:629–653
- Danaher J, McArthur N (2017) *Robot sex: social and ethical implications*. MIT Press, Cambridge, Massachusetts
- Danaher J (2019) The Philosophical Case for Robot Friendship. *J Posthuman Stud* 3(1) (Online first: unassigned to a volume or issue)
- De Graaf MA (2016) An ethical evaluation of human-robot relationships. *Int J Soc Robot* 8:589–598
- Dennis MJ (2019) Technologies of self-cultivation for contemporary life: how to improve stoic self-care apps. Jubilee Centre for Character and Virtue
- Devlin K (2018) *Turned on: science, sex and robots*. Bloomsbury, London
- Elder A (2014) Excellent online friendships: an Aristotelian defense of social media. *Ethics Inf Technol* 16(4):287–297
- Elder A (2019) *Friendship, robots, and social media: false friends and second selves*. Routledge, London
- Floridi L (2014) *The fourth revolution*. Oxford University Press, Oxford
- Hakli R (2014) Social robots and social interaction. In: Seibt J, Nørskov M, Hakli R (eds) *Sociable robots and the future of social relations*. IOS Press, Amsterdam, pp 105–115
- King O (2019) The good of today depends not on the good of tomorrow: a constraint of theories of well-being. *Philos Stud* (Online first: unassigned to a volume or issue)
- Kliniewicz M (2019) Robotic nudges for moral improvement through stoic practice. *Techné Res Philos Technol* (Online first: unassigned to a volume or issue)
- Lakatos G, Gácsi M, Konok V, Brúder I, Bereczky B, Korondi P, Miklósi Á (2015) Emotion Attribution to a non-humanoid robot in different social situations. *PLoS ONE* 10(3):1–28
- Mataric MJ (2017) Socially assistive robotics: human augmentation versus automation. *Sci Robot* 2(4):1–2
- Meghdari A, Minoo A (2018) Recent advances in social and cognitive robotics and imminent ethical challenges. In: *Proceedings of the 10th international RAIS conference on social sciences and humanities*
- Moravec H (1988) *Mind children: the future of robot and human intelligence*. Harvard University Press, Massachusetts
- Mori M (1970) The uncanny valley. *Energy* 7(4):33–35
- Mori M, MacDorman K, Kageki N (2012) The uncanny valley from the field. *IEEE Robot Autom Mag* 19(2):98–100
- Novikova J, Watts L (2014) A design model of emotional body expressions in non-humanoid robots. *Proceedings of the 2nd international conference on human-agent interaction*, pp. 353–360
- Peeters A, Haselag P (2019) Designing virtuous sex robots. *Int J Soc Robot* (Online first: unassigned to a volume or issue)
- Raj M, Semwal VB, Nandi GC (2018) Bidirectional association of joint angle trajectories for humanoid locomotion: the restricted Boltzmann machine approach. *Neural Comput Appl* 30:1747–1755
- Rossi S, Staffa M, Tamburro A (2018) Socially assistive robot for providing recommendations: comparing a humanoid robot with a mobile application. *Int J Soc Robot* 10(2):265–278
- Shaw-Garlock G (2009) Looking forward to sociable robots. *Int J Soc Robot* 1:249–260
- Skorupski J (1998) *Morality and ethics*. In: Craig E (ed) *Routledge encyclopaedia of philosophy*. Routledge, New York

- Snow N (2019) 'Virtue proliferation: a clear and present danger? In: Grimi E (ed) *Virtue ethics: retrospect and prospect*. Springer, New York, pp 177–196
- Sparrow R, Sparrow L (2006) In the hands of machines? The future of aged care. *Minds Mach* 16:141–161
- Taylor C (1991) *The ethics of authenticity*. Harvard University Press, Cambridge, Massachusetts
- Vallor S (2011) Carebots and caregivers: sustaining the ethical ideal of care in the 21st century. *Philos Technol* 24(3):251–268
- Vallor S (2012a) Flourishing on Facebook: virtue friendship and new social media. *Ethics Inf Technol* 14(3):185–199
- Vallor S (2012b) New social media and the virtues. In: Brey P, Briggie A, Spence E (eds) *The good life in a technological age*. Routledge, London, pp 193–202
- Vallor S (2016) *Technology and the virtues: a philosophical guide to a future worth wanting*. Oxford University Press, Oxford
- Van de Poel I (2012) Can we design for well-being? In: Brey P, Briggie A, Spence E (eds) *The good life in a technological age*. Routledge, New York, pp 295–306
- Van Grunsven J, Van Wynsberghe A (2019) Semblance of aliveness: how the peculiar embodiment of sex robots will matter. *Techné Res Philos Technol*, Online first
- Verbeek P (2012) On hubris and hybrids: ascesis and the ethics of technology. In: Brey P, Briggie A, Spence E (eds) *The good life in a technological age*. Routledge, London, pp 260–271
- Williams B (1993) *Morality: an introduction to ethics*. Oxford University Press, Oxford
- Wong P-H (2016) Responsible innovation for decent nonliberal peoples: a dilemma? *J Responsib Innov* 3(2):154–168

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