The Ethics of Artificially Intelligent Sexbots: A Philosophical Investigation into the Ethical Conditions for Human-Sexbot Interaction

by

Cindy Friedman

A dissertation submitted in fulfilment of the requirements for the degree

Masters in Philosophy

in the Department of Philosophy at the

UNIVERSITY OF PRETORIA

FACULTY OF HUMANITIES

SUPERVISOR: Prof Emma Ruttkamp-Bloem

August 2020

Acknowledgement

First and foremost, I dedicate this dissertation to my mother. While compiling this final draft of my work, I left this section until very last as my words kept failing me every time I sat down to write why I am so thankful to you and the role you've played in this journey. With every attempt, however, I realised that my words would never not fail me, and this bears testament to the fact that I don't think I could ever truly encapsulate how thankful I am for you. Your love, guidance, and unwavering belief in me throughout my life's journey has made this one possible.

I would like to thank my supervisor, Professor Emma Ruttkamp-Bloem. Your mentorship goes above and beyond your expert academic guidance and feedback. I count myself truly lucky to have had you as a supervisor – one who has taken a vested interest in me, my research, and my aspirations. You have had a profound impact on me and my research interests, and for that I will be forever grateful.

I would like to thank my father for reminding me how important it is to dedicate our time to those things we love the most. As any student and academic knows, research journeys are not ones filled with only love for what we do – the going gets tough sometimes and we may begin to question why we even started the journey in the first place. Your words of wisdom, however, were a constant reminder during those tough times to reflect upon my chosen path – one I chose because of my love of wisdom, and ardent belief in the real-world value of philosophy.

To my Tiaan, you were my sunshine on my most cloudy, rainy days. Your love and support know no bounds. Thank you for being my person.

Abstract

Discussion surrounding the development of artificial intelligence (AI), and its moral and social impacts, is one of the most important conversations of our time. Research within the ethics of AI is vital in ensuring that AI is developed and utilised in such a way that it remains beneficial to society. AI technology is utilised in various ways. This dissertation, however, focuses upon the use of AI technology in the form of sex robots ('sexbots'). Although research on sexbots within the context of AI and human-robot interaction (HRI) is gaining momentum, there is space for much more work to be done. Where a lot of focus has been on the moral and social impact of sexbots upon society, there is a gap in the literature as far as putting forward ways in which we may deal with these impacts is concerned – particularly preventing negative impacts, so that sexbot technology may remain beneficial to society. As such, my research enters this relatively new debate by way of not only drawing attention to the moral and social ramifications of sexbots, but also investigating ethical conditions for our interaction with them that may ensure sexbot technology is beneficial to society.

As such, this dissertation specifically conducts a ***philosophical investigation into the ethical conditions for human-sexbot interaction.*** It does so by way of investigating the moral and social problematics that may arise in relation to our interacting with sexbots in terms of *what* we may use them for (i.e. the various roles they may fulfill in society), as well as *how* we may use them (i.e. how we interact with them). Given the moral problematics that are discussed, the dissertation puts forward possible ethical conditions for human-sexbot interaction that we may consider as a way to try and ensure that sexbot technology is mainly beneficial to us as human interactants.

Specifically, this dissertation emphasises that when we consider the moral and social impacts of sexbots, as well as possible ethical conditions in relation to our interaction with them so as to prevent potential moral and social harms, we must do so from an *anthropocentric perspective* i.e. focus on *how* we use sexbots, as opposed to *what* we use them for. This is because we should not deem sexbot technology itself to be essentially 'good' or 'bad'. It is never technology itself that is good, bad, or neutral, but rather *how we use it* that deems it so. As such, *we* are the masters of our own moral fates in the context of ensuring that sexbot technology is beneficial to society.

Keywords: Artificial Intelligence, Ethics of Artificial Intelligence, Human-Robot Interaction, Sex Robots, Social Robots.

List of Contents

Introduction

In what has become somewhat of a holy grail in the field of AI, *Artificial Intelligence: A Modern Approach,* Russel and Norvig (2009: 1) begin their discussion of their definition of AI beautifully by stating:

> "We call ourselves *Homo sapiens* – man the wise – because our intelligence is so important to us. For thousands of years, we have tried to understand *how we think*; that is, how a mere handful of matter can perceive, understand, predict, and manipulate a world far larger and more complicated than itself. The field of artificial intelligence, or AI, goes further still: it attempts not to just understand but also to *build* intelligent entities".

The field of AI encompasses a variety of subfields, but it ultimately can be defined by organising it into four broad categories: creating computers that can think humanly, act humanly, think rationally, and act rationally (ibid.: 2). As such, one can succinctly define AI as "the subfield of Computer Science devoted to developing programs that enable computers to display behaviour that can (broadly) be characterized as intelligent" (Thomason 2016).

Although the foundations of AI can be traced back as far as Aristotle who was "the first to formulate a precise set of laws governing the rational part of the mind" (Russel & Norvig 2009: 5) by way of developing an "informal system of syllogisms for proper reasoning, which in principle allowed one to generate conclusions mechanically, given initial premises" (ibid.), its gestation can be accredited to Warren McCulloch and Walter Pitts (1943) and its official birthplace occurred at Dartmouth in 1956 thanks to John McCarthy, together with Marvin Minsky and Herbert Simon, among others (ibid.: 16-17).

77 years have seen blindingly rapid development within the field and the pace is escalating as there is currently a race towards the development of sophisticated AI technologies. Technologies such as autonomous weapons systems, self-driving cars, healthcare robots, and sex robots – to name but a few – are currently controversially in the media spotlight. Notwithstanding such reports, unless one is directly involved with the research and production of such technologies, one may still feel detached from this race. However, what one often forgets, or does not perhaps realise, is how deeply engrained AI has become into our everyday lives. From Netflix to Siri, to banking and Uber, we often forget

that we are utilising AI on a regular basis and that it, therefore, also alters the way we live our lives on a regular basis.

The discussion surrounding the development of AI can be said to be one of the most important conversations of our time given the broad impact its development and utilisation will in future have – and already has – on society[1]. With such rapid development, the discussion on the timeline towards reaching superhuman AGI[2] is a hot topic. With sceptics holding that it will not happen for hundreds of years, those on the other side of the spectrum hold that it may very well happen in our lifetime. Regardless of the timeline, however, AI could either be the best, or the worst thing, to ever happen to humanity depending on the direction in which development continues, and how humans respond to such development. It is therefore important that AI remains beneficial to society[3] (see e.g. Future of Life Institute

---

[1] Think of the use of autonomous weapons systems in warfare, the consequences of job loss to artificial intelligence, and the use of AI to, for example, perform surgeries and act as court judges (Tegmark 2018: 101-102, 105-107).

[2] Superhuman AGI refers to Artificial General Intelligence, which is the ability for an AI to accomplish any cognitive task in such a way that it supersedes human intelligence (Tegmark 2018: 39).

[3] The importance of ensuring beneficial AI has resulted in multiple initiatives, organisations and reports that attempt to put forward suggestions as to how the goal of beneficial AI can be achieved. Examples include the 2016 One Hundred Year Study of AI (AI100) at Stanford (https://ai100stanford.edu/), the Future of Life Institute (http://futureoflife.org), and the Malicious AI Report (https://maliciousaireport.com) written by Brundage, et al (2018). See also: The Human Rights Council's resolution on 'The right to privacy in the digital age' (A/HRC/RES/42/15) adopted on 26 September 2019 (https://digitallibrary.un.org/record/3837297?ln=en#record-files-collapse-header); the UNESCO Recommendation on Science and Scientific Researchers (2017) (http://portal.unesco.org/en/ev.php-URL_ID=49455&URL_DO=DO_TOPIC&URL_SECTION=201.html); the UNESCO Internet Universality Indicators (2019) (https://unesdoc.unesco.org/ark:/48223/pf0000367617?posInSet=1&queryId=a46642b0-1893-4f04-9bfb-b832b0851652); the Report of the United Nations Secretary-General's High-level Panel on Digital Cooperation on 'The Age of Digital Interdependence' (2019) (https://www.un.org/en/pdfs/HLP%20on%20Digital%20Cooperation%20Report%20Executive%20Summary%20-%20ENG.pdf); the outcomes and reports of the ITU's AI for Good Global Summits from 2017, 2018, 2019 and 2020 (https://aiforgood.itu.int/reports/). There are also frameworks related to the ethics of AI compiled by intergovernmental organizations, such as: the relevant human rights and other legal instruments adopted by the Council of Europe, and the work of its Ad Hoc Committee on AI (CAHAI); the work of the European Union related to AI, and of the European Commission's High-Level Expert Group on AI, including the Ethics Guidelines for Trustworthy AI (2019); the OECD's Recommendation of the Council on AI (2019); the G20 AI Principles, drawn therefrom, and outlined in the G20 Ministerial Statement on Trade and Digital Economy (2019); the G7's Charlevoix Common Vision for the Future of AI (2018); and the work of the African Union's

2017 & Floridi, et al. 2018), and hence the importance of research within the field of ethical AI. Philosophers, therefore, need to take their place behind the steering wheel so as to ensure that AI technology can be developed and integrated into our lives in a beneficial and morally responsible way.

Before embarking upon the elaboration of the topic with which this dissertation grapples, the distinction between morality and ethics which I will adhere to must first be clearly defined. There are moral dilemmas in our lives that we experience at a first-order lived experience level, such as whether or not to jump a red light, to steal a loaf of bread, to lie to a friend, to cheat in a relationship, etc. These moral dilemmas are addressed on a meta-level via different ethical systems, such as via religion or spirituality, tradition or culture, or even political systems, for example. Crisp (2011b) states that "… a contrast is drawn between morality, construed narrowly as involving in particular the notion of obligation and the sentiment of blame, and ethics, understood more broadly as covering all sources of reasons for living in one way or another". As such, ethics "[seeks] to find general principles or explanations of morality" (Crisp 2011a & see e.g. Skorupski 1998 and Rosenstand 2000).

In the context of ethics of AI, however, we need to tread carefully. This is because in this context we can only address moral issues from the bottom up – as opposed to the top down via the utilisation of a particular school of ethics to address moral issues. The reason for this is that in the context of ethics of AI we are considering a domain in which humans interact with technology, as opposed to interacting with other humans. As a result, apart from facing moral issues on their own, philosophers also have to consider relevant policy

---

Working Group on AI. Also important to consider are many initiatives and frameworks related to the ethics of AI developed by the private sector, professional organizations, and non-governmental such as: the IEEE's Global Initiative on Ethics of Autonomous and Intelligent Systems and its work on Ethically Aligned Design (2017); the World Economic Forum's 'Global Technology Governance: A Multistakeholder Approach' (2019); the UNI Global Union's 'Top 10 Principles for Ethical Artificial Intelligence'; the Montréal Declaration for a Responsible Development of AI, an initiative by the University of Montréal (2018); the Harmonious Artificial Intelligence Principles (HAIP) (2018). Hagendorff (2020) also "analyzes 22 of the major AI ethics guidelines and issues recommendations on how to overcome the relative inefectiveness of these guidelines" such as the Beijing AI principles (https://www.baai.ac.cn/news/beijing-ai-principles-en.html), AI Now 2019 report (https://ainowinstitute.org/AI_Now_2019_Report.pdf), DeepMind ethics & society principles (https://deepmind.com/about/ethics-and-society) and Artifcial intelligence at Google: Our principles (https://mail.google.com/mail/u/0/#search/dissertation/FMfcgxwJXBxDFQGlGBwSjvtlwpHcGbjr?projector=1&messagePartId=0.7).

documents – be they generated by government or industry – as well as the safety or professional ethical codes in the domain of technology.

As such, the first step in the context of ethics of AI is to identify the moral issues that we may be faced with. The second step is then to develop a code of ethics, or a set of ethical principles, that may address these issues – this is not a fully blown ethics, but rather a kind of merging of policy, safety and ethical considerations, and there are many examples of these (see e.g. European Commission 2019; Floridi et al. 2018; Select Committee on Artificial Intelligence 2018). Such sets of principles are, however, subjective to the extent that they mostly align with the value systems of the community/company/cultural group generating or supporting such a code – a good example is MIT's moral machine project. This may then, from a philosophical point of view, lead to a mirror meta-level moral dilemma in the sense that we still do not know what the right thing to do is objectively, faced simply by a set of ethical principles. This brings us to the third step. The third step is to finally tackle the issue of settling the right thing to do, by deciding into which school of ethics the code would best fit when considering who should enforce it and how it should be enforced. In this dissertation, I am specifically focusing on step one and to a limited extent on step two, while step three would potentially be future work as it is beyond the scope of a Masters dissertation.

The use of AI technology is broad and multifaceted. One such use of AI is within the field of social robotics. According to Duffy, a social robot can be defined as "[a] physical entity embodied in a complex, dynamic, and social environment sufficiently empowered to behave in a manner conducive to its own goals and those of its community" (2003: 177-178). Or, according to Breazeal (2004), social robots are "robots designed to interact with people in a human-like way". In being able to exhibit human-like emotions, humans are able to socially communicate with social robots and are able to form social bonds with them (Kanda, et al. 2012; Whelan, et al. 2018).

Dziergwa, et al (2018: 163) hold that:

"[T]he growing importance of […] human-machine interaction and the corresponding decline in the intensity of direct personal contacts that we observe in the past years is an irreversible process. This is why a lot of effort is put into HRI (human-robot interaction) studies, where social robots play the roles of personal assistants, trainers, therapists or teachers".

Various prominent AI researchers discuss the ways in which social robots may morally, and socially, impact the lives of human interactants and society at large. Vallor (2011), Sparrow & Sparrow (2006) and Sharkey & Sharkey (2010) discuss the implication of the use of social robotics for aged care. Sharkey (2016) and Komatsubara et al. (2014) investigate the possibility and impacts of the utilisation of social robots in classrooms. Turkle (2006) looks at the impact of companion robots, such as those which can play the role of a pet, on human interactants ranging from children to the elderly. Levy (2007) and Danaher & McArthur (2017) look specifically at social robots who can fulfil the sexual desire of human interactants. Other authors such as Evans (2010) and Hauskeller (2017) question whether robot companionship would be as fulfilling as companionship with another human being, and Coeckelbergh (2010a) investigates under what conditions social robots could even become our companions.

It is clear that social robots can be utilised in various different ways, to fulfil various different roles in our society. My focus in this dissertation, however, is on the capacity for social robots to be sex robots (or 'sexbots'). These are social robots with which users will be able to have human-like relationships – both physically and emotionally. I focus on social robots in this capacity because it is technology that will be welcomed into an extremely intimate part of our lives. In this form, technology will surpass its status as a useful tool – sexbots will be more than just tools we can utilise to look after and teach our children, or take care of the sick or elderly within our society. They have the potential to take on the role as peer and intimate companion. It is for this reason that particularly unique moral issues may arise (and these will be investigated).

Moreover, from an academic perspective, I focus on sexbots because it is an easy contextual standpoint from where to begin invesitigating and understanding how social robots may impact our lives. Sexbots will have an appearance that is as realistically human-like as possible, and human interactants will be able to interact with them in a way similar to how they may interact with a human sexual partner (albeit in a more limited capacity, and the reasoning for this will be explained). Thus, our interaction with sexbots will be the closest to human-like contact, as opposed to our contact with other social robots that take on non-human-like forms such as social robots that can fulfil the role of a pet dog, for example. It thus makes it easier to consider the potential impact of social robots on our lives from this context (the context of sexbots), because they are easier to relate to.

It must be noted that sophisticated sexbots are yet to come into existence and, as such, my understanding as to what sexbots essentially are in terms of what they look like, their

x

social capacities, and their potential utilisation, is based on general future trends that are discussed in robot ethics literature. However, despite the fact that we are yet to see the creation of the type of sexbots to which this dissertation refers, development within the field of social robotics is rapid. Therefore, it is worthwhile considering the moral and social impacts of sexbots as their creation may not actually be far off. Moreover, regardless of how imminent their creation may be, it is the role of the AI ethicist to question the impact of potential new technologies so that if, or when, the day arrives that this technology may be created and integrated into society, we can be aware of the impacts that may arise, and thus be prepared for them.

Particularly as far as research on sexbots in the context of ethics of AI and HRI is concerned, this dissertation makes a valid entry into a relatively new debate. The integration of social robots in society has garnered a lot of attention in the HRI community. HRI research has primarily been focused on the betterment of society. "Research goals include: creating robots to improve the social skills of children with autism, provide health-care and companionship to the elderly and infirm, and improve human interactions with robots as a whole" (Wagner 2018: 53).

The emergence of sexbots in particular has piqued public interest far and wide. However, in 2016, at the time of writing their paper, Scheutz and Arnold had noticed that engagement with the topic of sex robots and sexual behaviour had mostly been "casual inquiries, striking anecdotes, or vivid examples of sex robots [that] often shape the news coverage of the issues at hand" (2016: 2) with little actual academic engagement as far as the HRI community is concerned. Although academic ethical enquiries into sexbots is increasingly gaining momentum (see e.g. Adshade 2017; Devlin 2018; Danaher et al. 2017; Hausekeller 2017; Nyholm & Frank 2019; Richardson 2017; Sharkey et al. 2017), there is space for much more work to be done in not only bringing to light the ways in which sexbots may impact human interactants and broader society, but putting forward ways in which we may deal with these impacts – particularly preventing negative impacts, so that sexbot technology may remain beneficial to society.

As such, my research enters this relatively new debate by way of not only drawing attention to the moral and social ramifications of sexbots as far as interpersonal relationships are concerned, but also investigating ethical conditions for our interaction with them. I feel that this will be a valid entry into the current debate since I found that most literature talks about the consequences of the utilisation of this technology, but does not grapple as much with the question of ethical conditions in the context of human-sexbot interaction. In this

way, my research will also fit into the debate regarding the betterment of society through the utilisation of social robots and, particularly, sexbots.

Therefore, given this contextual background, my dissertation will specifically conduct a *philosophical investigation into the ethical conditions for human-sexbot interaction*.

In Chapter 1, I will investigate the nature of sexbots so as to understand what sexbots essentially are, and the purpose for their creation. I will also investigate the nature of our (human interactants') interaction with them. This lays down the conceptual groundwork for the chapters that will follow.

In Chapter 2, I will investigate the positive and negative impacts of sexbots on human interactants depending on what we may use them for. Firstly, I put forward the important distinction between *what* we use sexbots for, and *how* we use sexbots. This conceptual difference is important to understand for the following chapters, as Chapter 2 will specifically grapple with the implications of *what* we may use sexbots for, and then Chapters 3, 4 and 5 will deal with different aspects of *how* we may use sexbots and the implications thereof. As far as the distinction is concerned (introduced in Chapter 2), the aspect of *what* we use them for pertains to a categorical understanding of sexbots i.e. the roles they may fulfil in society. The aspects of *how* we may use them, however, is one which focuses on the human interactant and their behaviour towards sexbots, regardless of *what* they may be used for.

In Chapter 2, therefore, I deal with the 'what' aspect of the utilisation of sexbots and look at the various roles that sexbots may fulfil in our society and how, in each aspect of their utilisation, both positive and negative impacts may arise. In doing so, I exemplify how versatile and special sexbot technology is, as well as exemplify how we cannot deem sexbot technology itself as all good or all bad. Thus, I put forward an argument that when we consider possible ethical conditions for human-sexbot interaction, we must do so from an *anthropocentric perspective*. This is because it is not what sexbots are, nor what they can be used for that ultimately determines whether they may be morally or socially detrimental to humans, but rather, *how* we (human interactants) relate to them or treat them, i.e. *how* we use them, that determines this. Thus, *we* – as human interactants – are responsible for how the technology impacts us, and any conditions considered must be considered in relation to this aspect of *how* we use them, as opposed to *what* we use them for.

Having grappled with *what* sexbots may be used for, and how this may impact human interactants, I then question in Chapter 3 whether there is a way to try and ensure that sexbot technology remains beneficial to society. I suggest that one way to do so, is to consider *how* we use sexbots, i.e. investigating how our treatment of them, or behaviour towards them, may

impact us. It is beneficial for my overall research aim to do so given that although *what* we may use them for also has negative implications, as is discussed in Chapter 2, there are both positive and negative implications in each aspect of their utilisation. As such, considering conditions in relation to *what* they are used for – such as restricting certain roles they may fulfil in society – would not only mean preventing negative implications, but positive implications too.

As such, in Chapter 3, I thus look at our treatment of sexbots in a 'negative' sense by way of questioning the moral and social impact of treating sexbots *immorally*. I argue that treating sexbots immorally, may negatively impact human interactants and society at large. I therefore suggest that we should consider ethical conditions – in the form of formulating and implementing ethical boundaries – in relation to our interaction with sexbots that would inhibit human interactants from treating sexbots immorally. I argue that these boundaries should be considered from an anthropocentric perspective in that sexbots do not necessarily 'deserve' moral consideration (as they are not aware of how they are treated), but that ethical boundaries should prevent human interactants from treating sexbots immorally so that the moral fibre of human interactants is protected; i.e. we (as human interactants) need protection from our own potential immoral actions towards sexbots.

In Chapter 4, given my argument in Chapter 3 that treating sexbots immorally may be detrimental for human interactants and the society in which they live, and we should consequently consider ethical conditions that protect human interactants from their own immorality, I consider whether there is a way in which we could practically prevent human interactants from treating sexbots immorally. I argue that a way in which to do so may be to consider granting sexbots negative rights in the sense of Isaiah Berlin's (1969) distinction between positive and negative liberty – not so as to protect sexbots themselves from immoral treatment, but rather to protect human interactants from their own immoral behaviour; i.e. the negative way in which they (human interactants) may be impacted should they treat sexbots immorally.

Lastly, Chapter 5 looks at the treatment of sexbots in a 'positive' sense by way of investigating the moral and social implications should human interactants treat their robots *morally well* in the sense of 'meaningfully' interacting with them, specifically in the context of 'loving' them in a romantic way. Although the notion of 'love' with a robot is a controversial one, I put forward that it is possible to love a robot, however, this love would be different in nature as compared to love as understood between two people. Given the possibility of experiencing 'love' with a sexbot, I consider the ways in which experiencing

romantic love with a sexbot may negatively impact human interactants and society and then put forward how both roboticists (as far as their robotic designs are concerned), as well as human interactants, have a responsible role to play in terms of whether we allow such negative impacts to come to fruition. However, it remains the case that it is ultimately human interactants who are responsible for the way in which they interact with sexbots and, therefore, are responsible for any moral or social harms that may arise. Thus, in this context, the importance of maintaining an anthropocentric perspective in terms of *how* human interactants utilise sexbot technology remains.

In the Conclusion, I briefly recap the arguments of each chapter, unpack my contribution to the field, and point the reader to further possible research.

Chapter 1

The nature of sexbots and human interaction with them

1.   Introduction

As was mentioned in the introduction of this dissertation, sexbots are especially unique technological creations. Their potential to become integrated into society in deeply intimate ways means that there are very real ways in which sexbots may morally impact individual human interactants, and society, if the utilisation of sexbot technology becomes widespread. However, before we can analyse the potential moral and social impacts of sexbots, and the ways in which we could deal with these impacts, we must first have a clear understanding of what is even meant by the term 'sexbot' and why it would even be possible for them to become integrated into society in deeply intimate ways.

As such, this chapter will provide a clear understanding of what sexbots are by way of discussing the sense in which sexbots are robots. Then, I will show how the interplay between their (limited) sociability, and human-like appearance, paves the way for human interactants to perceive sexbots as being more than 'just machines', thus creating the possibility for human interactants to relate to them the way in which they may relate to another human companion – or at least this is how it will be perceived on the part of the human interactant.

Then, having a clear understanding of the nature of sexbots and how we may interact with, and relate to, them, I bring up the topic of consciousness in relation to sexbots and human interactants. This is for two reasons: firstly, I wish to emphasise that despite the intractability of the consciousness debate – not only in relation to robots, but also as far as humans are concerned – we should not let this deter us from seriously investigating ethical conditions for HRI in the context of sexbots, as such robots may already be able to mimic consciousness well enough to have humans interact with them as if they were conscious, i.e. *perceive* them as being conscious. This has important moral ramifications that will be brought up in the chapters

that follow. Secondly, an understanding of perceived sexbot consciousness is important for the discussion of robot moral status that is discussed in Chapter 3.

Finally, I will then consider the nature and scope of our interaction with sexbots in the sense of whether it is even possible for a human interactant to have sex with a sexbot. This is an important consideration given that, after all, sex will be an important component of human interaction with them.

2.   What are sexbots?

2.1.   In what sense are sexbots robots?

The notion of a robot is no unfamiliar concept. In simple terms, Bekey (2012: 18) defines a robot as "a machine, situated in the world, that senses, thinks, and acts". We are seeing robots making an appearance in various facets of our lives (albeit perhaps more pervasive at present in first-world or high- and middle-income countries). Initially, being utilised in labour and service industries, their use has expanded to the military and security industries and, nowadays, they are also deployed in research and education, as well as in the medical and healthcare industries (see e.g. Royakkers & van Est 2015).

Robots being used in these ways are essentially high-tech tools that are increasingly making our lives easier – and initially the essential aim of introducing robots into our lives was to free us up from work that is 'dull, dirty and dangerous' and thus the focus was on labour and service industries and the military in terms of, for example, bomb defusing and rescue robots. However, the capacity for their usefulness as tools has not remained their only function.

Increasingly, the notion that robots could have a very real place in our lives as potential companions is being marketed – e.g. care robots deployed in healthcare and care for the aged, nanny robots, and sex robots. Examples include AIBO, Pepper, SAM and Buddy[4]. Robots being designed so as to fulfil companionship roles fall under the category of *social robots*. Social robots are robots with which humans are able to socially interact in a simulated

---

[4] AIBO is a robotic dog designed by Sony (Sony 2020). Pepper is a social humanoid robot that is able to recognize faces and human emotions. It is available for use in businesses and in schools (SoftBank Robotics n.d.). "SAM is a mobile telepresence platform that autonomously navigates from room to room in a nursing facility, allowing staff to check in on residents on a regular schedule" (Ackerman 2016). Buddy is an emotional companion robot for home use (Buddy n.d.).

human-like way, i.e. humans interact with the robots *as if* they were human within certain limits, while the robots simulate certain aspects of human-human interaction, albeit also under certain important limitations[5]. This (simulated and limited) sociability sets them apart from other types of robots in that they do not only perform a technical function, but have a social dimension to them, too. Going back to Bekey's (2012) quote above, companion robots are not just physically situated in our world, but also socially. As Hegel et al. (2009) state:

> "[A] social robot combines technical aspects as well as social aspects – but the social aspects are the core purpose of social robots. The robot is not a social robot *per se*, it needs specific communicative capabilities to become a social robot. First, it implies the robot to behave (function) socially within a context and second, it implies the robot to have an appearance (form) that explicitly expresses to be social in a specific respect to any user. From this point of view, a social robot contains a robot and a social interface".

Given this understanding of social robots, sexbots can definitionally be understood to be a type of social robot. Sexbots can be seen to be the 21st century version of their sex doll precursors. Where sex dolls were designed to merely fulfil a technical function – by way of their providing a way for users to satisfy physical needs – the development of sexbots will fulfil not only this technical role, but a social role too, allowing them to be seen not just as tools to be used to fulfil a specific function, but also as peers and companions with whom human interactants can 'bond'[6] both physically, socially and emotionally. Levy (2007) sees this to be a very real possibility, having dedicated his book, *Love and Sex with Robots: The Evolution of Human-Robot Relationships*, to discussing how prevalent the occurrence of 'experiencing love' with a robot will become.

The development of sophisticated sexbots is, however, still in its infancy. Abyss Creations, for example, are the creators of RealDoll which, although considered to be the

---

[5] This interaction is limited as compared to human-human interaction given that robots are not conscious, free agents and, thus, interact according to the way they have been programmed. They are unable to genuinely reciprocate human sentiments. There will be more detailed discussion concerning this in the rest of this chapter.

[6] I say 'bond' in quotation marks because, presently, any bond formed with a sexbot will be unidirectional in the sense that the human interactant may bond with the robot, yet the robot will not bond with the interactant due to their incapacity to genuinely feel and reciprocate human sentiments. Whether this will remain the case remains to be seen, although it does not seem likely to change in the near future to more than simulated (and always limited) bonding, if at all.

3

most advanced from of sexbot today, is still a relatively unconvincing partner. However, the implications of their development are currently being seriously researched in the academic sphere by researchers writing on the ethics of social robotics (see e.g. Danaher & McArthur 2017; Devlin 2018; Levy 2007 & 2015; Richardson 2015; Scheutz & Arnold 2016), robotics companies continue to spend money on their development in many countries throughout the world, and international conferences and public seminars are being held that grapple with the nature and experience of human-sexbot interaction (such as the International Congress on Love and Sex with Robots). The prominence of their development has even brought about the Campaign Against Sex Robots (CASR), spearheaded by Kathleen Richardson, which "models itself on the longer-standing Campaign to Stop Killer Robots"[7] and opposes their development (Danaher, et al. 2017: 47). As such, despite the development of sophisticated sexbots being in its infancy, it is clear that it is urgent to grapple with the morality of their possible integration into society so that we are prepared for the challenges to the nature and quality of our intimate lives that they may bring. This illustrates the complex nature of the discipline of ethics of AI in the sense that there are real and current ethical issues – such as moral accountability in relation to self-driving cars[8] or biases in facial recognition technology[9] – with which we must be concerned, as well as serious calls to grapple with potential future ethical problems such as the concern over the singularity[10], or robotic

---

[7] The Campaign to Stop Killer Robots "was organised to preemptively ban fully autonomous weapons" (Danaher, et al. 2017: 47).

[8] In 2018, a self-driving Uber car killed a pedestrian in Arizona. This raised questions regarding who should be held morally accountable for the pedestrian's death: the car, the driver, Uber, or the car's programmers? (Griggs & Wakabayashi 2018).

[9] Concerns have been raised about racist algorithms being used in the police force (the most well-known example is the NYPD's stop-and-frisk programme) to predict potential crime hotspots and to conduct risk assessments of individuals to predict if someone is likely to commit an offence or become a victim of crime. The technology has proved to be prejudiced as reliance on the technology has resulted in more black men being stopped and searched than white men, for instance. This is due to the programming using historical data which is racially skewed (e.g. BBC News 2019; Crawford 2017).

[10] The Singularity refers to an intelligence explosion in which machines will become more intelligent than humans by way of a technological explosion that leads to "ever-greater levels of intelligence, as each generation of machines creates more intelligent machines in turn" (Chalmers 2010). Although there is no definitive consensus on when this may or occur, or even if it may occur, prominent researchers such as David Chalmers and Nick Bostrom grapple with the topic given the potential existential threats to humans that the singularity may bring if it does come about.

warfare[11], or in this case, the nature and quality of human-sexbot interaction and its impact on what it means to be human and interact with our world.

So, what would constitute a sexbot as being sophisticated enough to seriously be considered a 'true' companion? Danaher (2017b: 4-5) holds that sexbots must meet three conditions:

> "1. Humanoid form: it is intended to represent (and is taken to represent) a human or human-like being in appearance; 2. Human-like movement/behaviour: it is intended to represent (and is taken to represent) a human or human-like being in its behaviours and movements; 3. Some degree of artificial intelligence: it is capable of interpreting and responding to information in its environment. This may be minimal (e.g. simple preprogrammed behavioural responses) or more sophisticated (e.g. human equivalent intelligence)".

It is the intricate interplay between their human-like appearance and their sociability (due to their being artificially intelligent) that places sexbots in a league of their own among other robots, not only as far as robotics and the ethics of AI are concerned, but also as far as sex technology is concerned.

As far as their physical appearance goes, a distinction must be made between a humanoid robot, and an android robot. Where a humanoid robot's physical appearance *resembles* a human, an android robot is designed so as to *realistically* mimic the appearance of a human being – they *actually* look like human beings. Thus, there is a difference in appearance between the humanoid robot Pepper (who has a white, appliance-like finish, an endearingly cute yet unrealistically human-like face, and does not walk, but rather glides around on a set of wheels) and the android robot Sophia (whose creators are attempting to make her appear as realistically human as possible through a realistic skin tone and facial features). As such, "human-like robots, [are] often referred to as androids in the robotics literature to distinguish them from mechanical looking humanoid robots" (MacDorman 2005). Given this understanding, sexbots fall within the category of android robots. It is vitally important to understand the kind of technology with which I will be dealing, since this

---

[11] Serious concerns have been raised about the potential development of Lethal Autonomous Weapons Systems that will be able to achieve military objectives without human supervision. Thus, potential moral issues such as attributing moral responsibility for human deaths are raised (see e.g. Sparrow 2011).

is foundational for the rest of this chapter, and the chapters that will follow, so let us consider some more issues related to the technological nature of sexbots.

2.2. Understanding the 'social' in social robotics

It is the specificity of being able to communicate (to a limited extent) and physically interact with sexbots in a human-like way that may make for interactants not only seeing these robots as tools, but also seeing them as potential peers and companions, and even lovers (Goodrich & Schultz 2007: 205). This interaction is, in a bounded sense, 'reciprocal' in the sense of sexbots actually responding (albeit in a restricted manner) to our social advances towards them. The point – and reason for concern – is that however limited the nature of the reciprocity, interactants will not have to project meaning onto them as they may have to do with sex dolls, for example, as the latter have a mere passive presence (Turkle 2006).

This 'social' aspect of sexbots makes them unique as compared to other forms of sex technology, even those that take on a human-like appearance – such as sex dolls. However, sex dolls have been around for a very long time and their lack of 'social intelligence' has not necessarily been a hindrance when it comes to users potentially treating them as peers or companions. In fact, there is a whole subculture of sex doll users called iDollators whose members see their dolls not just as sex toys, but actual life partners (Beck 2013). This, however, is an uncommon scenario given sex dolls' lack of interactive capacities. People who treat their sex dolls in such a way are seen as social outliers.

Our interaction with sexbots, on the other hand, due to our being able to socially interact with them (albeit in a limited way), and their own limited capacity to reciprocate, draws upon our fundamental relational capacities, thereby making it all the more likely, and perhaps less strange, if people tend to want to 'bond' with their sexbots in the way they may bond with other people. Although this may still seem a strange state of affairs now, Levy (2007) argues that it will become more common in the not so distant future as, after all, this kind of scenario is part of the motivation of creating sexbots in the first place. They are "relational artefact[s]" in that they are "explicitly designed to engage a user in a relationship" (Turkle 2007).

Since sexbots are designed and created to be artificial substitutes or complements to human-human sexual interactions[12] (Danaher 2017b: 5), it is plausible to argue that sexbots are essentially human simulacra. This being the case, Dautenhahn (2007) warns that if we are attempting to simulate human beings, the social aspect of human nature is a vital aspect to consider. She states that "social intelligence is a key ingredient of human intelligence and, as such, a candidate prerequisite for any artificially intelligent robot" (ibid.: 682). Or, in other words, "developing an intelligent robot means developing first a socially intelligent robot" (ibid.).

Dautenhahn (2007: 684), based on the work of Breazeal (2002, 2003), Fong et al. (2003) and her own (1998), elaborates upon how the definition and conceptual understanding of social robots may vary depending on their purpose and how and why they interact with people and the environment in which they are situated. Social robots can be:

1. *"Socially evocative*. Robots that rely on the human tendency to anthropomorphize and capitalize on feelings evoked when humans nurture, care [for] or [become involved] involve with their 'creation'" (Breazeal 2002, 2003), are socially evocative.

2. *"Socially situated*. Robots that are surrounded by a social environment which they perceive and react to [are socially situated]. Socially situated robots are able to distinguish between other social agents and various objects in the environment" (Fong et al. 2003).

3. *"Sociable*. Robots that proactively engage with humans in order to satisfy internal social aims (drives, emotions, etc.) [are sociable robots]. These robots require deep models of social cognition" (Breazeal 2002, 2003).

4. *"Socially intelligent*. Robots that show aspects of human-style social intelligence, based on possibly deep models of human cognition and social competence" (Dautenhahn 1998), are socially intelligent.

5. *"Socially interactive*. Robots for which social interaction plays a key role in peer-to-peer HRI [Human-Robot Interaction], different from other robots that

---

[12] Danaher (2017c) discusses how sexbots could be used to complement human-human sexual interactions in scenarios where one partner may have a higher sex drive than the other. The partner with the higher sex drive could use the sexbot, thereby taking pressure off the partner with the lower sex drive to fulfil their partner's desires.

involve 'conventional' HRI, such as those used in teleoperation scenarios" (Fong et al. 2003), are socially interactive.

Sexbots can be said to fulfil all five of these properties, thus making them incredibly versatile and sophisticated in comparison to other social robots such as those mentioned above. This versatility and sophistication thus indicates how special sexbots are in the sense that they can be utilised, and interacted with, in so many different ways, thus also impacting our lives in so many different ways – and these potential impacts will be the focus of discussion in the chapters that follow.

It is clear that the social aspect of sexbots will enhance the likelihood that people may consider them as potential companions. However, not only is there the potential for people to see a sexbot as *a* (in the sense of one among many others) companion, but the sociability of sexbots informed by machine learning may also personalise interaction in such a way that interactants may come to see their sexbots as being a companion that is *unique* as compared to no other. They may potentially view the sexbot as *their* sexbot and not just *a* sexbot which they have purchased the way in which one buys a cellphone that one has no qualms about replacing if necessary.

Arnold and Scheutz (2018) suggest a dual understanding of robots as both types and tokens when considering questions around 'personal' robot identity. As far as the type-token distinction is concerned, its origin dates back to medieval philosophical interest in the universal versus the particular, although, as Arnold and Scheutz (2018) note, it can be referenced in its explicit form to C.S. Peirce. The distinction constitutes of a 'type' being more abstract in the sense of it being "general in occurring across different instances" such as a stop sign being used in different places, or the word 'the' being used in different sentences. However, a token is particular in that it will refer to that specific stop sign, on that specific street, or that particular 'the' which is "composed on a particular line and page, with pixels or ink, in a physical site" (Peirce 1998: 488).

Given this distinction, a dual understanding of robots as being both types and tokens implies that although robots can be a 'type' in the sense that in order for them to be reliable, predictable and effective, each type of robot is designed according to the same specifications; they are also 'tokens' in that each social robot has the potential for a certain particularity due to social robots functioning as "mobile, interactive systems, sharing time and space with particular users and interactants" (Arnold & Scheutz 2018). As such, it is possible that

8

"interactions with a [specific particular] robot can elicit great emotional investment from people" (ibid.), thus a social robot has a particularly unique presence to its user, as a token.

Sexbots are not only designed to look as realistically human as possible, and in such a way that they have human-like mannerisms and behaviour – such as their capacity to talk to us and physically interact with us to some extent in the same way in which a human companion would – but also they learn from their owner's specific behaviour, and in that sense, they may be viewed as unique from the owner's perspective, as 'their' companion (or even partner), which could lead to ascribing a token identity to a particular robot. After all, "if social robots remain 'types', i.e. if all instances remain the same irrespective of the contexts and individual interactions (as is, for example, the case with vacuum cleaning robots), they will likely disappoint their users who are seeking to connect with them on a personal level" (Breazeal 2002 as cited in Arnold & Scheutz 2018). Machine learning however promises more, as it pertains to "the automated detection of meaningful patterns in data" (Shalev-Shwartz & Ben-David 2014: vii). Machine learning practice mimics, to some extent, the learning processes of human beings, acquiring knowledge and skills through learning from (personal) experience in the sense that "[m]achine learning tools are concerned with endowing programs with the ability to learn and adapt" (ibid.) from input that is available to them. Here lies the potential for sexbots being viewed as 'unique' companions since, through machine learning, they may have the capacity to learn their partner's likes and dislikes, and will be able to adapt their behaviour according to the relevant inputs received from their environment.

2.3. The ethics of the human-like/android form of sexbots

Generally, social robots need not necessarily be realistically human-like in their appearance, just so long as interactants are able to interact with them by way of their being able to exhibit social behaviours. They may exhibit social behaviours by recognising users, assisting them and at times engaging in conversation. Paro, for example, takes the form of a baby seal yet allows for interaction with its users. The Paro website states:

> "PARO has five kinds of sensors: tactile, light, audition, temperature, and posture sensors, with which it can perceive people and its environment. With the light sensor, PARO can recognize light and dark. He feels being stroked and beaten [with a] tactile sensor, or being held [with] the posture sensor.

PARO can also recognize the direction of voice and words such as its name, greetings, and praise with its audio sensor. PARO can learn to behave in a way that the user prefers, and to respond to its new name. For example, if you stroke it every time you touch it, PARO will remember your previous action and try to repeat that action to be stroked. If you hit it, PARO remembers its previous action and tries not to do that action. By interaction [sic] with people, PARO responds as if it is alive, moving its head and legs, making sounds, and showing your preferred behaviour. PARO also imitates the voice of a real baby harp seal" (Paro Robots 2014).

As such, we could question the necessity for sexbots to take on a human-like form. Afterall, as Danaher (2017b: 5) questions, if sexbots are meant for sexual stimulation and release, why do they have to have a human-like appearance? Many sex toys are available that take on an animal form, for instance, and they still provide users with the physical stimulation that they seek. The difference with sexbots, however, is that they are designed so as to mimic "the real thing" (ibid.). And just as the social aspects of sexbots allow for interactants to relate to them as possible peers and companions, their human-like appearance may enhance this interactivity as well as the experience of the human interactant. As stated in the previous section, it is precisely the possibility of interplay between their social ability and their human-like appearance that makes it so urgent for us to set ethical conditions for interacting with them.

Moreover, a human-like appearance is important in relation to sexbots because it allows for more natural human-robot interaction (HRI) to take place as it significantly enhances the social aspects of the robot. As stated by Kanda et al. (2004):

"We believe that a robot partner, ideally, would have a humanlike body. A robot with a humanlike body allows people to intuitively understand its gestures, which in turn causes people to behave unconsciously as if they were communicating with a human…We believe that in designing an interactive robot, its body should be based on the human body to produce the most effective communication".

Moreover, not only may a human-like appearance allow for more effective communication, but it may also engender feelings of familiarity and affection from interactants. One of the first roboticists to suggest this was Masahiro Mori, head of the

robotics department at Tokyo University. This view is also borne out of the understanding gained from a study that confirmed that a robot with a human-like appearance will "provide immediate and comprehensible social cues, thereby enhancing a human's perception of any interaction with the robot and making it easier for the human to engage with it socially" (Levy 2007: 159-160), thus echoing Kanda et al's (2004) quote above.

Although such commentary on the importance of human-like robots now holds more traction due to recent technological advances, it is interesting to note that there has always been a general perception that a human-like robot appearance elicits the most positive responses from humans. Incorporating this perception into robot design dates as far back as the eighteenth-century Japanese robot designers (Levy 2007: 297)[13]. It must be noted, however, that it is important for designers to ensure they avoid the *uncanny valley* phenomenon coined by Masahiro Mori (1970). Mori predicted that "as robots appear more human, they seem more familiar until a point is reached at which subtle imperfections create a sensation of strangeness" (MacDorman 2005: 106). Mori held that "[m]achines that appeared too lifelike would be unsettling or even frightening inasmuch as they resemble figures from nightmares or films about the living dead" (ibid). As such, he cautioned robot designers to create robots with a humanoid appearance, as opposed to an appearance with complete human-likeness so as to avoid the uncanny valley. This phenomenon is well known within the robotics community, yet it must be noted that solid evidence proving its existence is sparse and that "the relation between appearance and behaviour in creating a subjective impression of familiarity or human presence has not been well explored, nor how habituation affects that impression" (ibid.: 107). Moreover, the Robot Accommodation Process Theory (RAPT) puts forward that "exposure to robots over time will reduce aspects of uncanniness, even in humanlike robots that are recognizable as artificial life" (Carpenter 2017: 280).

Whether or not explicitly focusing on designing the best human-like android robots is an ethical way in which to go about thinking about companion robots in general, needs a lot of careful reflection. There is concern that designing realistically human-like social robots that draw upon, and manipulate the human tendency to anthropomorphise, is deceptive and, therefore, fundamentally immoral (see e.g. Borenstein & Pearson 2012; Sharkey & Sharkey 2010; Sparrow & Sparrow 2006; Sullins 2012). Grappling with this issue in detail, and in its entirety, would go beyond the confines of this dissertation. The issue of deception is,

---

[13] Karakuri mechanical tea-serving dolls, for example, had a human-like appearance and motors inside their bodies that created realistic movements.

however, raised in Chapter 2 briefly, and Chapter 5 discusses it in slightly more detail, whilst also providing a critique.

3. An anthropocentric investigation of the ethical quality of human-sexbot interaction

3.1. The consciousness debate

Due to the human-like appearance of sexbots, and our being able to socially interact with them (albeit in a limited capacity), this creates the potential for human interactants to anthropomorphise them, and, therefore, perceive them as being *more* than just machines. As such, human interactants may, in a limited way, relate to their sexbot the same way in which they may relate to a human sexual companion. In doing so, human interactants may project human-like qualities on them, such as consciousness, because human interactants would want to believe that they have a 'true' connection with these machines as it would make their 'bond' seem all the more realistic. This next section looks at the issue of robot consciousness from two perspectives: the perspective of robots *actually* being conscious, and the perspective of human interactants *perceiving* robots as being conscious. The topic of consciousness in relation to our interaction with sexbots is important in that it adds to our understanding of the nature of our interaction with sexbots, and also is foundational to the understanding of robot moral status, which will be discussed in more detail in Chapter 3.

3.1.1. The actuality of robot consciousness

Are robots conscious? Can they feel what it is like to be a robot? To be treated in a specific way by a human interactant? How will we know? This is the problem of other minds[14] in a more modern context. Just as the debate surrounding the question of human consciousness seems to be somewhat intractable due to our being unable to "resolve issues of consciousness entirely through objective measurement and analysis (science)" (Kurzweil 2005: 380), so too seems the current debate surrounding artificial consciousness to be intractable.

---

[14] The problem of other minds, most commonly held to have arisen within epistemology, grapples with the question of: "how do I know (or how can I justify the belief) that other beings exist who have thoughts, feelings and other mental attributes?" (Avramides 2019).

The debate over artificial consciousness is as divided as the human case, with the broadest, and most traditional competing theories of mind remaining dualism and physicalism. Dualism and physicalism attempt to solve the mind-body problem from differing perspectives. The mind-body problem pertains to trying to understand the causal relationship that exists between our minds and bodies or, in other words, questions how our mental properties relate to our physical properties. Where mental properties are those properties which are private to each individual – be it the way in which you subjectively feel pain inflicted upon you, the memory that is evoked when you smell a rose, or your desire for chocolate ice-cream – physical properties, on the other hand, are properties that those around you can easily observe – such as the way you flinch when feeling pain, the smile that spreads across your face after smelling a rose, or the way in which you walk to the ice cream parlour so as to fulfil your desire for chocolate ice-cream.

How do these physically observable properties relate to the private mental properties that constitute the seemingly 'inner world' that exists within each of us? Dualists and physicalists starkly contrast each other on the matter (Robinson 2017). Substance dualists hold that "the mental and the physical are both real and neither can be assimilated to the other" (ibid.). There is thus more to us than our mere physical properties on this view. There is something it is like to be us, to feel pain, to smell a rose, to eat chocolate ice-cream. There is this rich inner world within all of us that is separate from our physical properties and is unique to each individual. Physicalists (in the reductive physicalist sense) on the other hand, hold the view that our mental states are merely physical states. Consciousness is merely a product of our body's physical processes and, as such, there is nothing beyond the physical and our mind is simply our body. The two are one and the same. Non-reductive physicalists, in their turn, and most notably property dualists, while not ascribing to identity relations between mental and physical properties, view physical properties as primary with mental properties relating to them via supervenience or emergence.

In relation to the question of artificial consciousness, consider Daniel Dennett's (1997) statement that: "we [humans] are extraordinarily complex self-controlling, self-sustaining physical mechanisms, designed over the eons by natural selection, and operating according to the same well-understood principles that govern all the other physical processes in living things". If consciousness is merely a product of physical processes, then from this point of view, why can't robots be conscious? If consciousness is the product of physical processes that occur in the brain, then surely, we can artificially replicate these processes, thereby creating a robot that, too, replicates consciousness? Property dualists, such as David

13

Chalmers, however, may not agree to this view as they might see it as merely grappling with the "easy problems" of consciousness. According to Chalmers, the easy problems of consciousness pertain to explaining the following phenomena: "the ability to discriminate, categorize, and react to environmental stimuli; the reportability of mental states; the ability of a system to access its own internal states; the focus of attention; the deliberate control of behaviour; the difference between wakefulness and sleep" (Chalmers 1995). However, there is more to consciousness than this.

If we were to artificially replicate the human brain, we would merely be creating an AI that acts *as if* it is conscious. However, it is far from clear in philosophical circles that consciousness can be determined behaviouristically (see e.g. Chalmers 2002; Kirk & Carruthers 1992). As such, creating an AI that behaves the way in which a conscious human being does, does not necessarily constitute it as being conscious. There is something more to consciousness. There is something it is like for us to be us. This is what Chalmers has coined as "the hard problem" of consciousness (Chalmers 1995), which pertains to the problem of subjective experience – and more to the point, *why* we have such experiences. Thus, approaching consciousness from a non-reductive physicalist perspective (e.g. property dualism) makes it difficult to believe that we would be able to create artificial consciousness. How could we, if we do not even understand how our own phenomenal consciousness comes about?

Turner (2019: 147) adds to the discussion on artificial consciousness by taking a different approach. He states that:

> "[C]onsciousness as qualia can be broken down into three stages as follows:
> For an entity to be conscious, it must be capable of (i) sensing stimuli, (ii)
> perceiving sensations and (iii) having a sense of self, namely a conception of
> its own existence in space and time".

Both stages one and two, he states, are present in various instances of artificial intelligence (ibid.: 147-148). Stage three, however, remains elusive as we have been unable to definitively prove that AI is capable of having a sense of self. Having a sense of self means having self-awareness, and only with self-awareness can one be said to be phenomenally conscious. Zahavi & Parnas (1998) discuss how there is an inextricable link between self-awareness and phenomenal consciousness. They hold that self-awareness is more than just being aware that one is perceiving an object, realising that one has mental states, identifying

oneself in the mirror, or referring to oneself using a first-person pronoun. Self-awareness is also a case in which one is aware of one's experience of something, resulting in one's subjectivity being revealed to oneself (ibid.: 689). In this way, self-awareness can be understood as consciousness being aware of itself. It is "acquainted with an experience in its first-personal mode of givenness" (ibid.: 691), meaning that although everything around us is presented to us in the same way, we each experience what is given to us differently, thereby making every experience subjective. If stage three (having a sense of self) remains elusive in AI, then we cannot definitively say that – or if – AI is phenomenally conscious.

The issue of robot consiousness is thus a very contentious and complex issue. Let us consider how we can best deal with this debate.

3.1.2. The perception of robot consciousness

Given the contentiousness of the artificial consciousness debate, it would seem more productive in relation to this dissertation to find a way to circumnavigate the debate entirely. In fact, for most AI engineers, the (im)possibility of artificial consciousness is a non-issue (Bringsjord & Govindarajulu 2018). We could also question whether the actuality of robot consciousness is important – to the extent of slowing down discussions of ethical limitations of HRI – if sexbots are in any case able to mimic consciousness well enough for interactants to *perceive* them as being conscious the same way in which humans are. In other words, my view here is that the contentiousness of the debate on artificial consciousness should not make us lose focus of the urgency of investigating ethical conditions for HRI in the context of sexbots, as such robots may already be able to mimic consciousness well enough to have humans interact with them *as if* they were conscious.

As such, we should focus less on the *actuality* of robot consciousness, and more on the *perception* of robot consciousness – at least for my purposes. Therefore, what is crucial here, as Arnold and Scheutz (2018) state, is "not what a robot is *in esse* but its function with and impact on people".

So, let us set aside the debate surrounding the actuality of robot consciousness, and rather focus on the possibility that human interactants may perceive their robots as being phenomenally conscious – whether they really are or not. The potential for interactants to perceive their sexbots as being conscious stems from them being, as Turkle (2007) states, a "relational artefact" in that these robots are "explicitly designed to engage a user in a relationship". This is due to their human-like appearance and social behaviour – as previously

discussed. Their appearance and behaviour work hand in hand to facilitate interactions that are as realistic as possible. Due to their human-like appearance and the capacity for sexbots to socially interact with interactants, as mentioned already, there is a high possibility that interactants will anthropomorphise these robots. Developers of social robots are acutely aware of this and thus take advantage of this tendency. Ultimately, these robots are intended to be human simulacra.

I would like to home in on what Kanda et al. (2004) state: that a robot with a human-like body "causes people to behave unconsciously as if they were communicating with a human". The tendency to behave in such a way is brought about by the natural tendency that people have to anthropomorphise non-human entities or inanimate objects. The tendency to anthropomorphise is the result of one's trying to "rationalise an entity's behaviour in a given social environment" (Duffy 2003: 180). It is hypothesised that the tendency to anthropomorphise evolved because it "favoured cooperation among early humans" and improved human fitness by helping us "distinguish between friends and enemies, to rapidly recognize predators, and to establish alliances with other tribes" (Damiano & Dumouchel 2018). As such, it is an evolutionary trait inherent within us all.

However, in stating that interactants may anthropomorphise robots, this does not mean that they believe that these robots are *actually* human. Rather, it means that a human-like appearance may evoke feelings within the interactant such that they view and treat their sexbots *as if* they are alive (Levy 2007: 99). One can extrapolate that this may be the case from studies conducted with AIBO, a robotic dog, where Peter Kahn and his team stated: "We are not saying that AIBO owners believe literally that AIBO is alive, but rather that AIBO evokes feelings as if AIBO were alive" (ibid.). Batya Friedman also led a research team that conducted similar research about AIBO indicating that "these feelings arise because people actually want to perceive their AIBOs as real pets, and therefore they attribute doglike emotions to AIBO" (ibid.). Similarly, with sexbots, if interactants want to perceive their sexbots as real people – as this will enhance their relational experience with them – then interactants may attribute human-like emotions to them and treat them as if they are human.

This is no futuristic prediction. Studies have found that people do tend to apply social rules to the computers with which they interact (Broadbent 2017: 640; Moon & Nass 2000). Broadbent (2017) notes that evidence for the application of such social rules comes from several angles: "people use stereotypical social categories, such as gender, ethnicity, and in-group and out-group status, when interacting with computers". Furthermore, she notes that "people use overlearned social behaviours, such as politeness, in their interactions with

16

computers" (ibid.: 640). Studies have expanded to show similar results as far as interaction with robots is concerned (ibid.: 640-641). The more human-like something appears to be, the more likely we are to anthropomorphise it. As such, given their android appearance, it is no leap in logic to then argue that the tendency to anthropomorphise sexbots will likely be high.

Due to interactants also possibly wanting to perceive their sexbots as being real people because of a desire for human companionship, this implies that they may want to perceive their sexbots as being phenomenally conscious as well, since this may make their relational experience with them all the more realistic. This point may be made because the presence of phenomenal consciousness plays such an important role in human-human relationships. People typically care about how their partners feel – not caring is seen as being cold or unsociable. Boltuć (2017: 224) states that "many people care about their partner's feelings, and the feel of love especially – those first-person private states. If one cares about one's partner deeply, then one does not just care about the partner's behaviour but what one's significant other feels for himself or herself" – we care about "what it is like for [our partners] to feel the experiences we share"(ibid.) with them.

Although love need not be present in order for an intimate encounter to be enjoyable, Boltuć (2017) holds that for many people, love is an important aspect of a relationship. If one cares about the giving and receiving of love in a relationship, then it is likely that one will care about one's partner's first person feel. If this is not the case, then "one may be inclined to care only about the behaviour, the performance of one's lover" (ibid.: 224). As such, the presence of phenomenal consciousness, this feeling of 'what-it-is-like' is an important aspect of human-human emotional bonds. After all, I am specifically speaking about human-like sexbots with which there will be the potential to not only have physical interactions, but social and emotional interactions too (even if one-sided). We are speaking of robots with which we have the potential to have human-robot emotional 'bonds'.

As far as human-robot relationships are concerned, interactants may then want to believe that their robot companions are phenomenally conscious, and not just behaving as if they are, because this will make their companionship with them seem all the more realistic. The social behaviour and human-like appearance of sexbots may engender such a reaction on the part of interactants. Although sexbots may, however, be 'philosophical zombies' incapable of genuinely reciprocating human sentiments in the sense that they may not have "felt emotions" (Vannoy 1980) but may instead display only "functional emotions" (ibid.), their human-like appearance and display of functional emotions may, however, mean that interactants may be able to have what seems to be a very realistic 'relationship' with them.

17

Interactants may very well want to view and treat their sexbot companions as if they are phenomenally conscious, and not just behaving as if they are, because this will make their companionship with them seem deeper and more realistically human-like – closer to having an actual human companion. Therefore, since there is a very real possibility that interactants may view their sexbots as being conscious, the actuality of their consciousness becomes a non-issue in the context of the arguments I make throughout this dissertation.

Now, having a better understanding of the nature of our interaction with sexbots, it is also worthwhile to analyse the specific concept of having sex with robots, as this is a vital aspect of the type of interaction that human interactants will be able to have with sexbots. Given the categorical distinction between "doing something to somebody, and doing something with [somebody]" (Migotti & Wyatt 2017: 24) and relating this to the act of sex such that there is a difference between performing a sexual act *upon* somebody, and performing a sexual act *with* somebody, we must consider whether it is even possible for a human interactant to have sex *with* a sexbot.

4. Can we have sex with sexbots?

In the field of ethics of AI, we often grapple with philosophical issues that we are yet to reconcile in relation to the human case. For example, as discussed in section 3, we are considering the possibility of artificial consciousness, despite not having very many real answers to questions surrounding human consciousness. We are considering how we could create an artificial moral agent, while we still grapple with questions surrounding human morality. Similarly, although the concept of sex is a definitionally grey area even in relation to human-human interaction, we now must consider whether we can have sex with robots, and what having sex with robots would entail, despite having no definitive answers to what sex entails as far as human-human interaction is concerned. We could all too easily simply state that it is possible for a person to have sex with anything, really – whether what is at the receiving end of this sexual conduct is designed for sexual stimulation or not. However, such a stance of the concept of sex would be problematic in our context given that I have argued that sexbots are more than mere objects in their role as social robots.

First, let us consider what it means to have sex with another person, and whether such an understanding could then be applied to interaction with a sexbot. Firstly, as noted above, a categorical distinction must be made between "doing something to somebody, and doing something with [somebody]" (Migotti & Wyatt 2017: 24). Relating this to sex, there is then a

difference between performing a sexual act upon somebody, and performing a sexual act with somebody. The distinction between the two depends on whether there is a shared sense of agency inherent within the sexual engagement. As Migotti and Wyatt (2017: 19) state: "having sex needs to be understood as a distinctive exercise of shared sexual agency" and that "agency is shared when people do things together with others, as opposed to simply alongside them". As such, there exists a "sexual we" (ibid.: 20) performing a sexual act together. In this instance, partners would be having sex with each other. Both participants are willingly engaging (physically and emotionally).

As far as doing something to somebody is concerned or, in the case of sex, performing a sexual act upon somebody is concerned, a partner may perform a sexual act upon someone where the other partner is not a willing participant. This may be the case in the instance of rape, where the act of sex is forced upon somebody. This may also be the case where one partner coerces the other into having sex, where the other partner is unwilling, but merely allows the sex act to occur without fully conscious engagement. There is a disconnect. In instances such is these, there is no "sexual we" (ibid.) performing the act. Participants are not having sex with each other, nor is one partner having sex with the other. One partner is simply performing a sexual act upon the other.

Such an understanding of having sex with somebody, calls into question the possibility of having sex with a robot. One could simply say that sexbots lack aspects of personhood that would enable them to be genuine sexual agents that can willingly engage in the sex act. Could we then say that it is not possible for an interactant to have sex with a robot – that they (the interactants) merely perform the sex act upon the sexbot? From one perspective, yes, since, as I have said, they cannot be genuine sexual agents. An interactant and a sexbot cannot genuinely have sex with each other.

However, only taking this aspect into account is superficial, given the intent behind the creation of sexbots and the way in which interactants view and relate to sexbots (as has previously been discussed in this chapter). As such, one cannot categorise sexbots along with the multitudes of sex toys available on the market, for instance. Referring to statements made by Matt McMullen (CEO of RealDoll), Migotti and Wyatt (2017: 24) state:

> "[D]esigners of robots are not in general aiming to create genuine agents, but
> instead to create the illusion of agency. They aim to create robots that are
> sufficiently humanlike for us to take the intentional stance towards them, i.e.

to attribute beliefs, desires, and experiences to them, and to develop affection towards them".

As such, lacking agency, sexbots cannot genuinely have sex with us. However, due to their having the "appearance of exercising the sort of agency required for someone to be our sexual partner" (ibid.: 23), interactants can attribute agency to them, thus creating the possibility for it to seem as if the robot is having sex with the interactant, while the interactant is having sex with it. Thus, creating a perceived "sexual we" (Migotti and Wyatt 2017: 20) on the part of the interactant. This relates back to the discussion about unidirectional relationships with robots. As such, the interactant can have sex with a robot, but this is simply due to the mimicking capacity of the technology itself.

Although one could say the same of a human-human sexual encounter – perhaps one participant is not genuinely engaging, but is merely pretending to do so – I take for granted that we cannot compare the two scenarios, given that humans have the actual capacity to be agents given that they are conscious, whereas robots are not. A robot, at least for now or in the near future, is unable to be a genuinely willing sexual participant.

5.   Conclusion

Now, having a clear understanding of the nature of sexbots, as well as the nature of our interaction with them, we can begin to consider how they may morally and socially impact us, as well as consider possible ethical conditions for our interaction with them. Chapter 2 will look at the positive and negative impacts of the various ways in which sexbots may be used. Before delving into this, however, I will introduce the conceptual distinction between *what* we use sexbots for and *how* we use sexbots. This distinction is important for the chapters to follow, and is also important in relation to my argument that we must use an anthropocentric perspective when investigating the possible moral and social implications of sexbots, as well as possible conditions for our interaction with them.

The next chapter specifically introduces this argument and then will go on to investigate the positive and negative impacts that sexbots may have on human interactants in the context of *what* they can be used for. This exemplifies how sophisticated and versatile sexbot technology may be as compared to other types of social robots (as was mentioned in section 2.2.). Moreover, addressing both the positive and negative ways in which sexbots may impact human interactants, and society at large, also emphasises the point made in

section 3.1. that we cannot let the actuality of robot consciousness hinder us from grappling with the very real impacts that sexbots may have on society, and the consideration of how to deal with these impacts. Their perceived consciousness is enough to make them socially situated in our lives in an extremely intimate way – a way which may have very real moral and social implications with which we must grapple. Also, and very importantly, the exemplification of the different ways in which we can use sexbots, and how there are positive and negative ramifications in each aspect of their utilisation, bolsters my argument for maintaining an anthropocentric perspective. My reasoning for why this is the case will be elaborated upon in the next chapter.

Chapter 2

The positive and negative impacts of *what* we may use sexbots for

1.   Introduction

An important distinction to note before continuing with this chapter is the difference between *what* sexbots are used for and *how* sexbots are used. One commonly conflates these two phrases: what is my laptop being used for? It is being used to write my dissertation. How is my laptop being used? It is being used to write my dissertation. However, there is a nuanced difference between these two phrases, which I think is important in relation to the interaction with sexbots – especially as far as our moral treatment of them is concerned (and this is discussed in Chapters 3 to 5). Concerning the '*what'* sexbots are being used for, this would be categorising them into different roles they can fulfill in society: such as carebots, companions or sex workers - as will be discussed here. It does not really touch upon aspects of our interaction with them and treatment of them but pertains to what they categorically are: what they are designed and manufactured to do. However, when one considers *how* sexbots may be used, this brings to light the aspect of our interaction with them and treatment of them. For example, we may make use of a sexbot as a companion with whom we share our homes and our personal lives – this is *what* they are used for. However, I may treat my sexbot immorally by way of using it to enact rape fantasies, or as an outlet for my anger by way of beating it or verbally abusing it after a bad day at work – this is *how* the sexbot is used. As such, *what* they are used for divides human interaction with them into various potential categories of use and the impact of that use on their interactants. However, *how* they are used is dependent upon the behavior of the human interactant towards the sexbot, regardless of the role that the sexbot may fulfill. It focuses on our *treatment* of them and how that reflects back onto us.

Given this distinction, this chapter will specifically be looking at *what* sexbots may be used for. I will consider positive implications and negative implications of each potential aspect of their utilisation. Not only does this provide a clear picture of the different ways in which sexbots could be integrated into society – and this is important to keep in mind for the chapters to follow – but, moreover, it demonstrates that the versatility of sexbots means that

we cannot essentially deem sexbots to be all good or all bad. The topic of the utilisation of sexbots is a nuanced one, which this chapter will demonstrate.

Moreover, it will also introduce the notion that given the nuances surrounding the utilisation of sexbots, and given that we cannot deem *what* they may be used for as all good or all bad, we should rather focus on the human interactants and how they interact with sexbots i.e. *how* they use them, and consider that it is them who are ultimately responsible for the moral 'goodness' or 'badness' of sexbots in terms of potential moral and social harms to society – not the technology itself[15]. This aspect of investigating *how* we interact with sexbots and the moral and social ramifications thereof will be dealt with in Chapters 3 to 5. This chapter, as was stated, specifically deals with *what* sexbots may be used for, and the moral ramifications thereof.

The following section will put forward my argument for why maintaining an anthropocentric perspective in relation to the consideration of ethical conditions of human-sexbot interaction is important. Thereafter, I will discuss the positive impacts of *what* we may potentially use sexbots for in the context of using them as companions for the sick and/or elderly, for sexual therapy for both sexual offenders and victims, as more satisfactory companions (as compared to other people) and as replacements for humans sex workers. I will then go on to consider the negative impacts of utilising sexbots in the same contexts as mentioned above.

2. Technological neutrality: the importance of an anthropocentric lens

Referencing Kranzberg's First Law (1986: 545), Whitby (2012: 237) states: "[T]echnology is never of itself good, or bad, or neutral. It is always all three, depending on how we use it". It is with this notion in mind that I will flesh out the ways in which sexbot technology may have positive impacts upon interactants and society, as well as ways in which sexbot technology may have negative impacts upon interactants and society. As such, this chapter can be seen as a *potential* cost-benefit analysis of the impact sexbot technology may have upon interactants and society as a whole. I say *potential* because much of the current literature on the topic draws conclusions based upon speculation by relating it to the impact that similar technology has had on society – such as video games and pornography – given that we are yet to see

---

[15] Roboticists also have some responsibility in this regard, but this will be discussed in more detail in Chapter 5.

sophisticated sexbot technology that is widely available (see e.g. Danaher 2017b; Sparrow 2017; Strikwerda 2017). As Danaher (2017b: 4) acknowledges in the introductory chapter to *Robot Sex: Social and Ethical Implications*, although some authors do take strong stances in relation to sexbot technology, all contributions to the book are speculative in nature and are "written in the shadow of an uncertain future". Section 3 of the chapter is a 'benefit analysis', where section 4 is a 'cost analysis'. Also, as mentioned often now, this is just one side of the story of the 'goodness' or 'badness' of sexbots, as the actual settlement in terms of cost or benefit is only finally determinable once we have considered also the '*how'* of sexbot use Chapters 3-5).

Before embarking upon this cost-benefit analysis, it must be noted that I wish to avoid the debate surrounding whether sexbots themselves should or should not exist. This is an extremely polarised debate. Where advocates of the technology, such as David Levy, push strongly for their development given his views that sexbots can not only enhance sexual experience, but can potentially be very real companions, others, such as Kathleen Richardson, advocate strongly against their development. Richardson has actually put forward a campaign called *The Campaign Against Sex Robots* (CASR) – also mentioned in Chapter 1 – modelled on the campaign to stop killer robots. Such strong advocation for or against their existence is however problematic, since both sides argue from the perspective of sexbots being inherently good or bad, without acknowledgement that it is essentially how we use and interact with these robots, and the subsequent impact of this utilisation and interaction, that may make them good or bad for human interactants and society at large. It is important to emphasise that it is *us* (the interactants) who are the focus here – not the robots. Following Whitby (2012), it is *how* we use them that will determine whether the technology is beneficial or costly to society, not the fact that they exist and have certain functions.

The introduction of new technology into society will always carry with it both benefits and risks. Drawing on work done by Ibo van de Poel, Danaher (2017c: 120-122) compares sexbots to the introduction of the iPhone in 2007. At the time of its release, there would have been speculation about its potential positive and negative impacts on society, but it is only now, 13 years later and looking in retrospect, that we know the full extent of its effects. Despite not knowing the extent to which the technology would impact society both positively and negatively, the technology was still released without insistence that Steve Jobs – or society or academia – give it more thought and scrutiny.

Why the scrutiny about sexbots then? I can understand both sides. We need to scrutinise the technology given its intimate nature and consequent impact on a core aspect of

24

who we are, yet constant scrutinisation may hinder its development, and then we may never know the potential positive impacts this technology may have had on society. As such, in this chapter, I scrutinise the potential positive and negative effects but not so as to advocate for or against the integration of this technology into society or not, but rather so that if, or when, they are integrated, we can at least try to minimise negative effects on humans, while enhancing positive effects. This is very much in line with the stances of well-known contributors to the ethics of AI, such as Asaro (2012), Bostrom & Yudkowski (2014) and Wallach & Allen (2009).

As such I here argue that we consider what and how sexbots can be used for from an anthropocentric perspective in the sense of suggesting that we should analyse the impacts of the technology through the lens of how people may interact and use them, and not through the lens of analysing whether the technology itself is inherently good or bad. Although there are aspects of sexbot technology in which we may analyse the technology itself, such as with sexbots that take on the form of children[16], even this can, controversially, be analysed to be both good and bad depending on how and why the technology is used.

Such an anthropocentric lens is important to keep in mind in relation to many aspects of technological development. Increased connectivity to the internet, through smart phones for example, is beneficial in the sense that it offers access to increased knowledge, but it also depends on how users seek this knowledge and further disperse it. The spread of fake news, and consciousness surrounding its spread, has been on the rise in recent years. There is a tendency to blame technology for this: it is the fault of social media, internet bots, or the devices that make access to fake news so easy. However, blaming technology for the problem of fake news is incredibly short sighted. The main problem is actually *us*:

> "People are the principal culprits, according to a new study examining the flow of stories on Twitter. And people, the study's authors also say, prefer false news. As a result, false news travels faster, farther and deeper through the social network than true news. The researchers, from the Massachusetts Institute of Technology, found that those patterns applied to every subject they studied, not only politics and urban legends, but also business, science and technology. False claims were 70 percent more likely than the truth to be shared on Twitter. True stories were rarely retweeted by more than 1,000

---

[16] This will be discussed in more detail in section 4.2.2.

people, but the top 1 percent of false stories were routinely shared by 1,000 to 100,000 people. And it took true stories about six times as long as false ones to reach 1,500 people. Software robots can accelerate the spread of false stories. But the M.I.T. researchers, using software to identify and weed out bots, found that with or without the bots, the results were essentially the same" (Lohr 2018).

Similarly, as far as the positive and negative impacts of sexbots are concerned, my focus shifts away from blaming the existence of the technology itself for any positive or negative impacts upon society, and shifts towards the interactants utilising the technology. My assumption will be that it is how we use sexbot technology that determines their moral and social impact on society (this will be the focus of my discussion in Chapters 3 to 5)

However, as promised, before turning focus towards *how* sexbots are used, I will firstly, in the sections that follow, look at the positive and negative aspects in relation to each aspect of *what* we may use sexbots for. Doing so allows me to demonstrate why considering conditions in relation to what we may use them for is problematic, thus allowing us to, in the following chapters, see why it is better to focus on *how* we use sexbots as far as investigating ethical conditions for our interaction with them are concerned. As such, I will consider the general arguments within the debate on sexbots and what they can be used for. In each aspect of their use I will draw attention to both the benefits and drawbacks of that particular aspect of their use.

3.   Positive impacts

There are various ways in which sexbots could be used that may positively impact human interactants. Such instances which will be specifically discussed include: companions for the disabled and/or the elderly (especially in care facilities), sex therapy, being more satisfactory partners and/or making human companionship more satisfactory, and replacing prostitutes. In each instance, only the positive impact of their utilisation will be discussed. Any problems that may arise with such utilisation will be reserved for discussion in the Section 4.

3.1. Companionship for the disabled and/or elderly

Although not a necessity, sex and sexuality are "important elements in a fulfilled life such that their nonvoluntary absence from someone's life would be morally relevant" (Di Nucci 2017: 75). In fact, sex and sexuality play such an important role in people's lives that "[s]exual liberty has increasingly come to be regarded as a basic human right" (Appel 2010: 152). The organisation *Health, Empowerment, Rights and Accountability* (2004) states that:

> "Sexual rights are a fundamental element of human rights. They encompass the right to experience pleasurable sexuality, which is essential in and of itself and, at the same time, is a fundamental vehicle of communication and love between people. Sexual rights include the right to liberty and autonomy in the responsible exercise of sexuality".

When referring to sexual rights, however, it is important to distinguish between positive sexual rights, and negative sexual rights[17]. Broadly speaking, as far as positive rights are concerned, it pertains to having the right to something such as the right to healthcare or the right to education. Negative rights, however, are rights that specifically "call for the prohibition of some action or the right not to be interfered with" (Hirschl 2000: 1071). They are rights that entitle us to freedom from or to something. Examples include the right to freedom of speech, or the right to freely practice your own religion. In relation to sexual rights, then, positive sexual rights pertain to the right to be provided with sexual pleasure whereas negative sexual rights pertain to "the right not to be constrained or interfered with in [our] sexual attitudes and behaviours" (Di Nucci: 2011).

Di Nucci (2011) holds that positive and negative sexual rights are incompatible. This is because:

> "[if] A has a positive sexual right, then that means that there is at least one person who would lack negative sexual rights. Namely: the person who would be supposed to fulfil A's positive sexual rights. If everybody has negative sexual rights, then everybody has the right to refuse to fulfil A's sexual needs, but then A has no positive right to sexual pleasure" (ibid.: 159).

---

[17] This distinction is based on the distinction between positive and negative liberty, first presented by Isaiah Berlin (1969) and will be discussed in detail in Chapter 4.

More than this, it also seems arbitrary to equate sexual positive rights to other positive rights such as the right to food and water, or the right to an education. Whereas positive rights such as these play a fundamental role in our survival and wellbeing, sex, on the other hand, cannot be likened to such necessities. Governments surely are not responsible for ensuring that all citizens are provided with satisfactory sexual gratification (see e.g. Appel 2010 & Di Nucci 2011)? However, we can at least acknowledge that nobody should be hindered from enjoying sexual fulfilment (within certain moral limits), thus focusing on negative sexual rights.

Groups of people who may experience nonvoluntary absence of sex from their lives include the severely mentally or physically disabled, and the elderly who suffer from neurodegenerative diseases (Di Nucci 2017: 73), especially if such individuals reside in care facilities. As Appel (2010: 152) notes, some disabled individuals may have physical limitations so extreme that they are incapable of self-stimulation. However, even if they were capable of self-stimulation, "the act of physical contact with another human being appears to provide a degree of pleasure and fulfilment entirely distinct from mere masturbation, even when that contact lacks emotional intimacy" (ibid.). Moreover, capabilities aside, residing in care facilities also "constrain[s] the ability of persons to express their sexuality and sexual needs" (Tarzia, et al. 2012: 609). This is due to many factors such as institutional policies and regulations, privacy, and attitudes of the staff (ibid.).

Particularly in relation to older patients suffering from dementia, residing in residential aged care facilities, Tarzia et al. (2012: 609) note that "aged care facilities do not tend to be environments that are conducive to, or supportive of, the freedom of sexual expression". This despite the fact that sexuality and intimacy remain important elements of people's lives later in life, and also that, specifically in relation to dementia, "[s]exual sensations are among the last of the pleasure-giving biological processes to deteriorate, and are an enduring source of gratification at a time when pleasures are becoming fewer and fewer" (Roach 2004).

Central to debates surrounding sexual rights and the disabled, however, are issues of consent and vulnerability. As far as consent is concerned, there is concern that those who are mentally disabled may not have the "the ability to give or withhold consent to the formation of a relationship and physical intimacy" (Tarzia et al. 2012: 610), hence there is uncertainty regarding their willingness to partake (Di Nucci 2017: 85). This raises concern surrounding the potential for abuse (ibid.). Yet, as Di Nucci (2017: 88) states:

"Both protecting patients' welfare and consent requirements are obviously paramount, even before getting into the issue of sex robots; but we must be careful that, in exercising prudence about patients' welfare and consent requirements, we do not lose sight of what healthcare exists for, namely, care – and in this case sexual care. So, for example, denying sexual care to all patients with mental/learning disability may end up erring on the side of caution".

As such, denying the disabled sexual gratification stifles their autonomy and personhood – and can thus be seen as a failure of duty to care (Tarzia et al. 2012: 611).

Such issues, however, only arise in relation to human-human sexual interaction. It can be argued that the utilisation of sexbots is a way to overcome these problems. As far as consent is concerned, "sexual activity with someone who is unable to consent is a sexual offence" (Bartlett 2012: 143). Both human parties need to have the capacity to consent to sexual relations. However, this issue does not arise in relation to human-robot relations since, in the context of using sexbots for the disabled, the human interactant could initiate the interaction, and the sexbot would then respond appropriately. There would be no forward advances on the part of the sexbot that could be deemed inappropriate, or result in abuse given that these robots can be "programmed so as to minimise, if not eliminate, the risk of abuse" (Di Nucci 2017: 85). As such, sexbots are a potential way to protect the negative sexual rights of the elderly and/or disabled in a safe way.

More than providing sexual pleasure, sexbots could also provide care and companionship for the elderly and/or disabled in the form of caregiving practices. As far as the utilisation of sexbots in relation to the sick or elderly is concerned, although an aspect of their utilisation could be the possibility for them to provide sexual companionship, it could also be considered that these android robots could not only provide companionship of an intimate nature, but could also fulfil a type of caregiving role, thus bridging the gap between carebots and sexbots. As such, these android (sex) robots could also be a type of carebot by way of taking care of the sick and elderly by performing menial tasks such as keeping track of the cared-for's vital signs, reminding them to take medication or eat a meal, or alert loved ones or relevant authorities when something goes wrong. An additional aspect of such care could be them also providing intimate sexual companionship for elderly human interactants.

This is a relevant consideration given that – as discussed above – sexuality and intimacy remain important elements of people's lives later in life and that giving the sick or

elderly freedom to enjoy sexual gratification can be seen as an important aspect of the duty to care. Yet, this is not an aspect of caregiving that human caregivers could be expected to provide since doing so would infringe upon their negative sexual rights. However, the utilisation of sexbots – that could also perform other basic caregiving duties – would ensure that those being cared for can enjoy their positive sexual rights, without any person having their negative sexual rights infringed upon. Such companionship could be seen as an aspect of duty to care that robots will have the capability to potentially fulfil.

According to Vallor (2011: 252):

> "Carebots are robots designed for use in home, hospital, or other settings to assist in, support, or provide care for sick, disabled, young, elderly, or otherwise vulnerable persons. The kind of support they may provide varies widely, but generally following Sharkey and Sharkey (2010), we can distinguish between the following actual or potential functions of carebots: performing or providing assistance in caregiving tasks; monitoring the health or behavioral status of those receiving care or the provision of care by caregivers; and providing companionship to those under care".

Such carebots are currently being designed. An example is the Actron MentorBot™ created by US-based robotics corporation Acrotek (Sparrow & Sparrow 2006: 2) and, more recently Robear – a robot that can "transfer frail patients from a wheelchair to a bed or a bath" (Hurst 2018). Although it had been advertised that the Actron MentorBot™ could also act as a companion, as mentioned above, the importance of sexual intimacy for the disabled and elderly is generally overlooked and, as such, it seems that intimate companionship has not been considered in relation to carebots. However, given that there is rapid technological development in the realm of sex robotics, as well as care robotics, I suggest that the prospect for the two realms to converge is a pragmatic possibility. As Sparrow & Sparrow (2006: 5) note: "[g]iven that robotics technology is advancing steadily, it would be perilous to found any judgement about what might be possible in the future solely on the basis of a detailed analysis of existing technological systems".

3.2. Sexual therapy

3.2.1. Therapy for sex offenders and victims

The idea has been put forward that sexbots could be used in therapy to treat sex offenders such as rapists or paedophiles. Although some, such as Patrick Lin, feel that the very idea is dubious and repulsive (see e.g. Knapton 2017), there have been strong arguments made in its defence with Ronald Arkin claiming in 2014 at a robotics conference that "robots designed to look and act like children could be used to treat pedophiles the way methadone is used to treat drug addicts" (Strikwerda 2017: 133). By this it is meant that drug addicts are treated for their addiction by way of not cutting them off from substances completely, but by providing them with "[m]edications developed to treat opioid use disorders [that] work through the same opioid receptors as the addictive drug, but are safer and less likely to produce the harmful behaviors that characterize a substance use disorder" (National Institute of Drug Abuse 2018). Similarly, then, the utilisation of child sex robots could be a way to treat sex offenders by way of not cutting them off from enacting upon any sexual tendencies, but providing a safer space for them to 'heal' or act upon their impulses in such a way that will be less harmful to society, or unethical.

Although the very act of enacting a rape scene upon a sexbot, or engaging in sexual activity with a childlike sexbot may itself seem repulsive – reasons why will be elaborated upon in section 4 – the idea does warrant attention given that there are positive impacts that may come about with such utilisation and the intent would overall not be immoral as this would happen in a context of therapy. Strikwerda (2017) compares the utilisation of computer-generated child pornography to treat paedophiles to child sexbots. The idea behind computer generated child pornography is that it provides a safe outlet for antisocial feelings or tendencies that may otherwise be directed towards actual children (ibid.: 138). Although scientific evidence proving this is scarce, Dutch sexologists Van Beek and Van Lunsen make such claims by referring to a study by Diamond el al. (2010) "who found that the number of reported cases of child sex abuse dropped markedly when the production, distribution, and possession of child pornography was decriminalized in the Czech Republic for a while" (Strikwerda 2017: 138). Of course, in this instance, the decriminalisation of child pornography – in which actual children play a role – is highly problematic given that children would be victimised and harmed, while in the case of computer-generated child pornography that is not necessarily the case.

"Ordinary child pornography is considered harmful to children. In acts of sexual contact between adults and children, mutual consent is in general assumed to be absent, and, therefore, they are always considered sexual abuse. Since the production of images showing a minor engaged in sexually explicit conduct typically involves an act of sexual contact between adults and children, such images can be considered recordings of harmful sexual abuse. Following this argument, not only the production, but also the possession and distribution of child pornographic images can be considered harmful to children. After all, the possession and distribution (i.e., consumption) of child pornography supports the market for it, and, therefore, causes the (obviously extremely harmful) sexual abuse of children inherent in the production of such materials. Moreover, the child pornographic images themselves are a permanent record of the sexual abuse that occurred in the production, which could continue to haunt the children concerned when they grow up, especially if discovered by others" (ibid.: 136-137).

"[L]egal philosophers commonly acknowledge that the moral limits of criminal law are determined by the harm principle" (ibid.: 135) whereby "power can be rightfully exercised over any member of a civilised community, against his will, [if the aim] is to prevent harm to others" (Mill 1986: 6). Since, in this instance, children would be harmed, child pornography must remain criminalized so at to protect children from harm.

However, given Strikwerda's (2017) comparison to computer-generated child pornography, we could consider the utilisation of sexbots instead of decriminalising child pornography since, currently, sexbots cannot be said to be harmed given that they are not conscious and cannot *feel what it is like* to be harmed were they to be sexually abused given them lacking phenomenal consciousness (as was discussed in Chapter 1). As Strikwerda (2017: 136) states in relation to computer-generated child pornography:

"[E]ntirely computer-generated child pornographic images do not involve a child really engaged in sexually explicit conduct. They consist of computer-simulated depictions and not of recordings of child sexual abuse. Therefore, the production, distribution, and possession thereof cannot be considered harmful to children, and hence the harm principle does not apply".

The same can be said about child sexbots given that these robots lack the capacity to be harmed since they are not phenomenally conscious – they cannot *feel what it is like* to be harmed in this way. As such, there exists the possibility that sexbots could be used to treat paedophiles, or sex offenders in general, by way of providing a safer outlet for them to indulge in their antisocial behaviour. If it is potentially the case that even child pornography – which is not interactive at all – may decrease the likelihood of child sexual abuse, then perhaps an even more interactive experience with a sexbot could prove to be even more beneficial?

On the other hand, there has also been mention of not only using sexbots to treat sexual offenders, but also to treat the victims of sexual abuse. As McArthur (2017: 41) states: "… people who have experienced sexual trauma often find it difficult to form intimate relationships. Sexbots might help people overcome such trauma through sexual experiences that are safe and controlled". As such, sexbots could provide a safe space for both offenders and victims to work through psychological problems and traumas.

3.2.2. Therapy for sexual dysfunction

As explained above, not only is there potential for sexbots to be used therapeutically for sexual offenders, but there is also potential for them to be used therapeutically for victims of sexual abuse. However, another therapeutic use may be for people who may generally suffer from what Levy (2007: 308) terms as "psychosexual hang-ups". In fact, for Levy, the therapeutic use of sexbots in this instance is inevitable (ibid. 300). One specific way in which sexbots could positively help interactants is through their replacing human sex surrogates:

"Surrogate therapy is a three-way process, with many of the sessions involving the client, the surrogate partner, and the patient's therapist. It is the therapist who decides when the client is ready to work directly with the surrogate on their emotional and sexual problems, who introduces the client (usually male) to the surrogate (with the therapist present), and who consults with the surrogate when the therapist feels that the client is ready for intimate and private contact with her. And while the client is attending sessions with the surrogate, he is still being counselled by his therapist, who is also in regular contact with the surrogate. The treatment process is designed to develop the client's skills at physical and emotional intimacy. All of the most

common sexual dysfunctions and their causes can be treated by surrogate-partner therapy, including premature ejaculation, nonconsummation of a relationship, erection difficulties, performance anxiety, and fear of intimacy. The surrogate and the client usually progress through a series of structured exercises in relaxation, introspection, communication, nurturing, and sensual and sexual touching" (ibid.: 216).

There is therefore real potential for sex surrogacy to be a successful method of treatment. For example, The School of Intimacy, Consciousness and Self-Awareness in the United Kingdom has achieved a high degree of success by using the method, noting that their "Sexual Surrogate Partner Programme has enabled hundreds of men and women to overcome sexual dysfunctions, gain control of their sexual responses and lead full and satisfying sex lives" (The School of ICASA n.d.). However, despite there being a professional association (The International Professional Surrogates Association), it has never really taken off in a big way. This is possibly due to the associated controversy of paying for sex, as well as "complex legal, moral, ethical, professional and clinical applications" (Noonan 2003). For example, there was concern from therapists in the 1980s about "recommending the use of surrogates to their clients in case of an eventual legal action should the client contract a sexually transmitted disease in the process" (Levy 2007: 218).

Given these concerns that have hindered the potential for sex surrogate therapy to be used more widely, sexbots could be employed as sex surrogates as opposed to humans, being programmed with the necessary "psychosexual knowledge, teaching skills, and humanlike sensitivity" (Levy 2007: 219). In this way the positive effects of the therapy could be retained without potentially harming any humans or infringing on any human rights.

3.3. More satisfactory companionship

Levy, in various instances in his book *Love and Sex with Robots: The Evolution of Human-Robot Relationships* (2007), suggests that sexbots could provide more satisfactory sexual companionship than human companions ever could. This is mainly due to his standpoint that we will be able to customise our robot partners in such a way that we will be able to make our perfect partner, thus providing far greater satisfaction than a human companion ever could. Evans (2010) refers to this argument as the "greater satisfaction thesis" (GST) which is the

thesis that "[i]t is likely that romantic relationships with robot companions will be more satisfying than romantic relationships with human companions".

Levy's belief is not one that may only apply to a few exceptional people – such as iDollators who have romantic relationships with their sex dolls[18]. Rather, he holds that more satisfactory relationships with robots may become a common occurrence – as Evans (2010) states: "[it] is plausible to interpret Levy as claiming, then, that *many* people will find romantic relationships with robots more satisfying than romantic relationships with human beings" due to their customisability – not only physically, but behaviourally too. For example, "people who are worried about infidelity will be able to request robot companions programmed to be completely faithful. Alternatively, those who prefer the frisson of uncertainty may specify that their robot companion be programmed to have some probability of being unfaithful" (ibid.). Evans (2010) refers to this as the total specification argument – although he fundamentally disagrees with Levy's (2007) argument.

Not only is there the possibility that sexbots may be a more satisfactory partner for a single human interactant themselves, but there is also the possibility that the addition of a sexbot within a human-human romantic relationship may make the relationship between the two people more satisfactory, although in an indirect way. Sexbots could be a solution to desire discrepancy problems, as well as tension around the kind of sex that partners may desire.

As far as desire discrepancy is concerned, this is the instance, especially in longer term relationships, where one partner desires sex more frequently than the other, thus creating tension within the relationship on the part of the partner who does not feel satisfied within their relationship, as well as on the part of the other partner who feels pressure to keep up with their partner's physical desires. The sexbot could then be an outlet for the partner with higher sexual desire, and release pressure of the low-desire partner. This will help with guilt and resentment, allowing the partners to then have sex with each other that is "mutually desired and mutually fulfilling" (McArthur 2017: 41).

The second instance in which sexbots could help partners within relationships is solving the problem of discrepancy between the kind of sex that partners want. Sexbots could, again, be an outlet for the partner that has sexual fantasies that their partner may be unwilling or too uncomfortable to fulfil. At the same time, the other partner would also feel less pressurised to satisfy the sexual fantasies of their partner.

---

[18] See Chapter 1.

35

It is suggested that the role of a sexbot in both these instances could help prevent infidelity within relationships. The reason is that although there are many causes for infidelity, "two of the leading ones are dissatisfaction with the amount of sex available in the relationship and the desire for novelty and variety in our sexual partners" (McArthur 2017: 42).

3.4. Replacement of sex workers

Levy (2015) has put forward the idea that sexbots could be beneficial by way of their potentially replacing human sex workers. This is not a far-fetched idea given that in both Tokyo and South Korea, sex doll escort services have become a lucrative business. As Levy (2015) states: "If static sex dolls can be hired out successfully, then sexbots with moving components seem certain to be even more successful".

Levy (2015) holds that (some of) the reasons why people – specifically men[19] – may seek the services of a human sex worker, could be fulfilled just as well by a robot. These reasons are variety, lack of complications and constraints, and lack of success with the opposite sex:

> *Variety:* Sex workers provide the opportunity for clients to have sex with a range of different women, as well as engage in various kinds of sexual practices. Sexbots could provide the same opportunity as far as appearance, personality or voice is concerned, as well as satisfying the sexual desires of clients who have/have had partners unwilling to accommodate their sexual preferences.
>
> *Lack of Complications and Constraints:* For many clients "the principal benefits of the commercial sex exchange include the clear purpose and bounded nature of the arrangement, as well as its anonymity, its brevity and the lack of emotional involvement" (ibid.). Sexbots will more than adequately satisfy these human motivations.

---

[19] "In contrast to the relatively well-researched topic of men paying for sex, there is almost no systematic published research on the reasons why women pay, or on what exactly they are seeking. But what little published evidence there is on this topic suggests that the reasons are close to those that motivate the male clients of sex workers, principally a lack of complications and constraints and a lack of success with men" (Levy 2015).

Lack of Success with the Opposite Sex: For many reasons, some men struggle to meet and develop relationships with women. This may be due to social anxiety, physical disablement, or unattractiveness. For these men, paying for sex is the only way in which sex may be available. Sexbots could easily fulfil this role given that they "will be immune to any ugliness or physical deformity in their clients, and to their clients' psychological inadequacies" (ibid.).

As far as the ethics of the utilisation of robot sex workers is concerned, Levy (2015) also raises three benefits: Firstly, there is negative stigma surrounding the hiring of human sex workers, and robot sex workers could mitigate this by way of the fact that they are not 'real people' but unconscious, unfeeling objects. Secondly, being able to freely make use of robot sex workers may mitigate certain relationship issues, similar to the ones mentioned in the subsection above. Here, Levy (2015) notes:

"…there will be many who feel that the sexual demands placed on them within their relationship are excessive, and who will therefore appreciate a night off now and then, in the knowledge that what is taking place is nothing 'worse' than a form of masturbation. There will also be some who positively relish the idea of robots, programmed to be sexually adept, teaching their partner to improve their lovemaking skills. And there will be couples, both of whom derive pleasure and sexual satisfaction from a threesome in which the third participant is willing to indulge in whatever sexual activity is asked of it (subject of course to its programming and engineering)".

Thirdly, there is a common conception that sex work is a degrading profession, one in which hard drugs and sexually transmitted diseases and infections are a part, and an industry that is fed by the human trafficking industry. Although not all agree with such a conception of sex work necessarily being a 'bad thing', replacing human sex workers with robot ones could mitigate some of these problematic aspects of sex work[20] (see e.g. Taylor 2017).

---

[20] One would think the possibility of sexbots reducing rates of human trafficking would have received more attention – whether it is a realistic probability or not. However, there has been very little discussion on the topic

4. Negative impacts

Having investigated how various ways in which sexbots may be used could bring about positive impacts for human interactants, I will now investigate how the very same ways they could potentially be used, could also bring about negative impacts.

4.1. Companionship for the sick and/or elderly

While the technological convergence of sexbots and carebots (discussed in section 3) has not taken place, but given the need for disabled and elderly care, and the arguments made for the importance of sexual intimacy, I evaluate the drawbacks of using sexbots/carebots to look after and be companions to the sick and/or elderly "based on a common sense or pragmatic assessment of the nature of the tasks that robots might be required to carry out" (Sparrow & Sparrow 2006: 5). Given this, and given the scarcity of literature that grapples with the question of sexbots as companions for the sick and/or the elderly, I draw more heavily on literature that deals with the drawbacks of carebots specifically being used for the sick and/or elderly in particular, and expand upon the arguments made so as to incorporate the sexual companionship aspect of sexbots/carebots.

When discussing the topic of using carebots for the care of the sick or the elderly, and concerns surrounding the benefits and the drawbacks thereof, it is important to acknowledge the difference between "concerns about the replacement of human care by AI assistive technologies and, on the other hand, concerns about care assisted by AI technologies (as opposed to care not assisted by such technologies) but without replacing human care" (Coeckelbergh 2010b: 181). Commonly, and Coeckelbergh (2010b: 181) acknowledges this, the former is far more controversial than the latter, and this is because it is commonly held that "care provided by AI systems is not as good as human care" (ibid.: 182) – that human care is irreplaceable because AI systems do not *really care*[21] for patients.

As far as using carebots to assist human caregivers is concerned, it is far less controversial in the sense that carebots could help with more menial tasks like those mentioned above, thus freeing up the time of human caregivers so that they can care more

---

and when conducting research on the topic, most sources discussing the matter are news websites (see e.g. Taylor 2017) or blogs (see e.g. Rising 2018).

[21] This is because they are (at least for now) not phenomenally conscious, and thus cannot experience caring feelings towards a person.

deeply for patients on a social and emotional level (see e.g. Coeckelbergh 2009; Sparrow & Sparrow 2006). Given the controversy surrounding the replacement of human caregivers with carebots, however, similar controversy arises if we consider the replacement of human caregivers with robots that will not only care for patients, but be legitimate companions.

I will thus focus this section particularly on the *replacement* of human caregivers, with particular emphasis on the companionship aspects of sexbots/carebots, and how this may be problematic not only for those being cared for (i.e. cared-fors), but also for those who are doing the caring (i.e. caretakers). It is also important to consider the ramifications of replacement given that, for example, Sparrow and Sparrow (2006: 12) hold that "regardless of the intentions of the designers and manufacturers, in reality robots will inevitably be used to replace human staff. The work pressure on aged-care staff, within both the community and residential sectors, and within both private-for-profit and not-for-profit aged-care homes, is already great".

4.1.1. Impact on cared-fors

The utilisation of sexbots in the role of carers and companions could be detrimental to those people being cared for by the robot. Three main issues can be raised in this regard: Firstly, the replacement of human caregivers with robots could increase social isolation by way of reducing or eradicating human contact; secondly, the introduction of robots to the sick/elderly as companions can be seen as inherently unethical and harmful to human interactants given that their potential happiness will be based on deception; thirdly, the replacement of human caregivers with robots could cause those being cared for to feel undignified (see e.g. Coeckelbergh 2010b; Sharkey & Sharkey 2010; Sparrow & Sparrow 2006).

Sparrow and Sparrow (2006) emphasise how important it is for cared-fors – and particularly those who are elderly – to have *human* social and emotional contact, and that lack of human contact is detrimental to their wellbeing, stating that "[t]here is a large body of research which shows that frequent interpersonal communication is critical for good aged care" (ibid.: 13) and that "[e]ven when there is impairment of cognitive capacity, the need for involvement with other people remains, and there is much evidence indicating the value of social interaction and appropriate communication, both verbal and non-verbal, when caring for people with dementia" (ibid.), for example.

In some instances, the elderly already have minimal social contact with their families or other support structures. Reducing this social contact even more by way of introducing

robot companions is, for Sparrow & Sparrow (2006), indefensible. Both Sharkey & Sharkey (2010) and Sparrow & Sparrow (2006) note that the replacement of human caregivers with robots may very well lead to neglect on the part of their families or support structures. As Sharkey and Sharkey (2010: 25) state: "There is a risk that … the provision of a robot companion could be used as justification for leaving the senior citizen on their own for longer. It could be used to alleviate the family's guilt about doing so – 'Don't worry about Granny, she's got the robot to talk to'".

Although it could be argued that some form of companionship is better than no companionship – and yes, this may be the case, and a robot companion would be more beneficial than someone spending their days in isolation – the cared-fors would still benefit more were they to be in a friendly social environment with other people (ibid.: 24). Also, although some companionship may be better than no companionship, it is argued that this is not 'real' companionship given that the relationship is non-reciprocal. The robot does not really 'care' for the patient, the way a human caregiver could. Although it is not always the case that human caregivers are necessarily caring, given the sometimes demanding nature of their jobs, human caregivers at least have the capacity to care, whereas robots – currently – do not (see e.g. Coeckelbergh 2010b).

This last point about robots not really being able to 'care' feeds well into the second problem that arises in relation to the replacement of human caregivers with robots: that the introduction of robots to the sick/elderly as companions can be seen as inherently unethical and harmful to human interactants given that their potential happiness will be based on deception. Both Coeckelbergh (2010b) and Sparrow & Sparrow (2006) touch upon this. As Sparrow & Sparrow (2006: 17) state: "… robots are clearly not capable of real friendship, love, or concern – only (perhaps) of their simulations". According to Sparrow & Sparrow (2006) then, these incapabilities on the part of robots have the following three implications: firstly, emotional work is an important aspect of caregiving which robots will be unable to fulfil; secondly, any benefits that cared-fors gain in being looked after by a robot will be as a result of being deluded about what robots are capable of offering; thirdly, it is "morally reprehensible" to even consider placing robots in such caregiving roles (ibid.).

Although there is nothing wrong with the happiness that robots may bring about through companionship, it is what brings about this happiness that Sparrow & Sparrow (2006) find problematic. This happiness will be a result of the delusion that human interactants have that their caregiving robot really does love and care for them and there are two reasons why this is problematic:

"Firstly, failure to apprehend the world accurately is itself a (minor) moral failure. We have a duty to see the world as it is. It is a sad thing to be deceived about the world; it is a bad thing to perpetuate and prolong such deception ourselves … Secondly, and more importantly, such deception is a bad thing because our preferences are unlikely to be met, our interests advanced, or our well-being served, by illusions. What most of us want out of life is to be loved and cared for, and to have friends and companions, not merely to believe that we are loved and cared for, and to believe that we have friends and companions, when in fact these beliefs are false" (ibid.: 17-18).

As such, given that replacing human caregivers with robots will only make those being cared for happy by way of deception, this is – if one follows this line of argument – not beneficial to those human interactants – there is no actual improvement in their wellbeing – and it is in fact detrimental to their wellbeing (ibid.: 18). This is happiness brought about by way of deception, which will then come to replace human contact which at least has the potential to bring about genuine happiness[22]. Moreover, as Turkle et al. (2006: 360) note, just the idea of a human interactant expressing their love and affection for a robot who says 'I love you' in return, creates a sense of uncomfortability; "it raises questions about the kind of authenticity we require of our technology".

Thirdly, there is concern that the replacement of human caregivers with robots could cause those being cared for to feel undignified. This is related to concern surrounding the deception of robot companionship in the sense that encouraging the elderly to bond with robot companions by way of deception can be seen as a way of infantilising them (see e.g. Sharkey & Sharkey 2010: 26). It also seems undignified in the sense that replacing human caregivers with robots may mean their families or support structures may visit less, making it seem as if those who require care are not worthy of human contact and concern – they can

---

[22] It must be noted that issues surrounding deception being unethical do not only exist in the context of the utilisation of social robots for taking care of the sick or elderly who may be suffering from neurodegenerative diseases (such as Alzheimer's disease or dementia) wherein it may be more likely that cared-fors could be deceived. Scholars such as Borenstein & Pearson (2012), Sharkey & Sharkey (2010), Sparrow & Sparrow (2006) and Sullins (2012) put forward that deception occurs, and is ethically dubious, in various instances wherein social robots may be utilised by various different kinds of human interactants – even those who have their full mental capacities. This issue will be discussed in more detail in Chapter 5.

be handed over into the care of a robot while families and support structures have a false sense of concern.

4.1.2. Impact on caregivers

Literature on care robotics most commonly focuses on the impact that the technology would have on those being cared for. However, Shannon Vallor in her article entitled *Carebots and Caregivers: Sustaining the Ethical Ideal of Care in the Twenty-First Century* (2011) brings to the fore the overlooked aspect of how care robotics may negatively impact not only the cared-fors, but the caregivers as well. Moreover, not only are caregivers generally overlooked in this regard, but when the caregiving practice is taken into consideration, it is cast mainly in a negative light by way of focusing on how caregiving is more of a burden on caregivers than anything else. As such, when considering the replacement of caregivers by robots, there is limited focus on how carebots may help to reduce the burden for caregivers and on society (Sharkey & Sharkey 2010), allow caregiving to be a choice rather than an obligation (Borenstein & Pearson 2010: 283-284), allow caregivers to trade menial tasks for emotionally meaningful ones, thus allowing caregivers to provide deeper care (Coeckelbergh 2010b: 183), and decrease the numbers of undercompensated and overworked caregivers (Borenstein & Pearson 2010: 285; Coeckelbergh 2010b: 183).

As such, literature mainly focuses on how replacing human caregivers with robots would mainly *benefit* the caregivers because caregiving is seen as being nothing but a burden and that "[e]ven the provision of companionship and emotional support is characterised as a 'task'" (Vallor 2011: 255). However, replacing human caregivers with robots cannot be seen as all good. There are potential drawbacks to the replacement of human caregivers for those caregivers themselves due to there being "good internal to the practice of caregiving that we might not wish to surrender, or that it might be unwise to surrender even if we might often wish to do so" (ibid.: 256).

Vallor (2011) uses three conceptual lenses to show that there is good internal to the practice of caregiving, and that surrendering these goods may be detrimental to caregivers themselves, and even society as a whole. The three conceptual lenses used are: virtue ethics, care ethics and the capabilities approach.

As far as virtue ethics is concerned, replacing caregivers with robots may result in caregivers not cultivating the virtues of reciprocity and empathy. With regard to reciprocity, although it is, according to Aristotle, a natural disposition (ibid.: 257), we must cultivate it as

42

a virtue so that we can "reciprocate *well*, in the right ways, at the right times, and as appropriate to particular circumstances and people" (ibid.: 257). Referencing Lawrence Becker (1986), Vallor (2011: 258) states that "reciprocity is in fact among the most fundamental virtues, one that is essential to creating and sustaining the primary goods that allow human lives to flourish". One such context in which reciprocity can be cultivated is in the practice of caregiving: in being there for others, we may learn that others will be there for us in our times of need. In this way, we see how "reciprocity holds human relations together" (ibid.). Moreover, it is not only a case of giving so that we may one day receive, but also giving because by doing so we develop our own moral character in that it facilitates the development of other virtues, such as "patience, empathy, and understanding" (ibid.). Thus, according to Vallor (2011), replacing caregivers with carebots would "[deprive] us of an important opportunity to cultivate reciprocity in ourselves, and to understand its centrality to human bonds" (ibid.: 258).

With regard to empathy, Vallor (2011: 258) states:

"Though Aristotle did not himself consider empathy a virtue (1984, NE 1106a4), I have claimed (2011) that it occupies a critical role in the life of virtue Aristotle describes and that Aristotle would likely himself have recognized it as a proper virtue had he enjoyed a better understanding of the integration of emotion and reason in human persons, as well as the essential nature of human vulnerability in grounding ethical human relations".

Genuinely empathising with someone is more than having just goodwill for them in the sense that "[f]or those with whom I empathize, I will indeed 'take trouble', and this occurs most notably in caring relations. In caregiving practices I take upon myself the trouble of recognizing, acknowledging, feeling, and responding appropriately to the suffering of another" (Vallor 2011: 259).

In replacing human caregivers with robots, not only will the care being given to those being cared-for lack genuine care and emotion, but this replacement will also prevent the virtue of empathy being cultivated. For, as mentioned in the previous subsection, replacing caregivers with robots may allow for more remote forms of care which will not allow caregivers to fully grapple with what the cared-for is experiencing, thus empathy cannot truly be cultivated. According to Vallor (2011: 269), if virtues such as reciprocity and empathy

(amongst others) are not cultivated, the "impact upon our moral character, and society, could be quite devastating".

Drawing on Noddings' (1984) work on care ethics, Vallor (2011) notes that the caring relation is one that comes naturally to all human beings and regulates our behaviour in the sense that it motivates us to "restore the relation when it becomes deficient or disrupted" (ibid.: 260). Our understanding of the caring relation is manifested through our memories of being cared-for throughout our lives. This caring relation is, however, not a moral one in and of itself, yet from it, an ethical ideal is acquired. In acquiring this ethical ideal, we acquire "a notion of the 'ethical self' that comes to function as a commitment to meet the other … from within the caring attitude" (ibid.). Drawing from Noddings (1984), Vallor (2011: 260) importantly notes that it is only through caring for others that one can ensure that one's "ethical self is sustained and enriched". In other words, "the ethical self as a regulative ideal" (ibid.) is developed through our memories of being cared for, but one can only come to care for that ideal through committing to caring for others, thus caring for the ethical self, and sustaining and enriching it.

Vallor (2011) notes two central criteria required for there to be a genuinely caring commitment towards another person: engrossment and motivational displacement. Where engrossment entails being orientated towards the reality of another person, as opposed to the reality that is my own, motivational displacement then entails the resultant feeling of this engrossment which is the feeling of needing to respond to the other person's reality, such as by reducing their pain, for example. These two criteria mean that one cannot care for another person remotely since this would lack engrossment in the other's reality. As such, replacement of a human caregiver with a robot would mean something would have been done to care for this person in need, but this caring would not sustain the caregiver's self as an ethical being (Vallor 2011: 260-261).

Therefore, concern is raised that the replacement of human caregivers with robots would liberate people from caring, thus seducing us into not caring at all. However, it is not necessarily the case that people will not care at all, but perhaps this care may be of a different nature: it may not be care "in the sort of intimate, direct relations that presuppose engrossment in the reality of the suffering or vulnerable other" (ibid.: 262). In order to protect ourselves from this possibility, it is suggested by Vallor (2011) that we must then "attend more carefully to the ways in which engagement in caring practices is critical for our own well-being, and not just that of those who need us" (ibid.).

44

Lastly, Vallor (2011) explores the issue through the lens of the capabilities approach. Vallor (2011) focuses on three capabilities that will not be developed if robots come to replace human caregivers: affiliation, practical reason and emotion.

As far as affiliation is concerned, it pertains to the ability of recognising and showing concern for other people (Nussbaum 2006: 77 & Vallor 2011: 263). Relating this to the caregiving practice, Vallor (2011: 263) states: "… it is through directly giving care that I learn to recognize in concrete others those expressions of need and desire that first motivate the attitude of 'concern' on my part".

Practical reason pertains to "[b]eing able to form a conception of the good and to engage in critical reflection about the planning of one's life" (Nussbaum 2006: 77). Caregiving can develop the capability of practical reason in two ways: firstly, caregivers gain full understanding of caring relations, what it entails, the inherent goodness of the caring role itself and what it means to care well (Vallor 2011: 264). Secondly, in caring for another person, caregivers become capable of truly planning their life in the sense that one cannot plan one's life well unless one is aware of "the vulnerabilities of sickness, injury, age, or other conditions of life" (ibid.) to which one will become exposed. In caring for another person who has to face such vulnerabilities, one becomes reflective about how one may cope with such vulnerabilities in one's own future.

Lastly, emotional capabilities pertain to: "Being able to have attachments to things and people outside ourselves; to love those who love and care for us, to grieve at their absence; in general, to love, to grieve, to experience longing, gratitude and justified anger. Not having one's emotional development blighted by fear and anxiety." (Nussbaum 2006: 76-77). According to Vallor (2011: 265), one can only develop emotional capability if one becomes deeply involved in the practice of caregiving in such a way that one comes face to face with the various emotions that caregiving entails – both positive and negative – such as fear and vulnerability at the prospect of losing the person to whom one has become attached through caring for them, or happiness and gratitude when the person one is caring for has a good day, such as someone suffering from dementia who has a lucid day, or an arthritic patient who has a less painful day. This develops the capacity for a "rich emotional life" (Vallor 2011: 265). One cannot develop this capacity through remote care should human caregivers be replaced with robots.

Regardless of the lens used to analyse the impact of replacing caregivers on the caregivers themselves, it is clear that it is the very 'caring' aspect of human care that allows caregivers to flourish as human beings – not the menial, habitual tasks of making sure

medication is taken, or food is provided. Should caregivers be replaced with robots that have intimate functionalities that can perform menial tasks, 'compassionately' care for patients, as well as be intimate companions, human caregivers will become redundant, and human care will no longer be needed, which may be detrimental to human caregivers and society at large.

Human caregivers may, essentially, become less caring people due to that very 'caring' aspect of caregiving being fulfilled by carebots/sexbots. This may not only be detrimental to caregivers who are formally employed in caregiving roles, but also to family members who may have taken on a caregiving role for an older relative. If, as mentioned in the previous section, family members can hand over care of an older relative to a carebot, they, too, are free to not really care at all. Although one could contend that the same may apply to family members who place older relatives in care facilities with human caregivers and never visit, at least in such an instance, cared-fors are being looked after by caregivers with the capacity to care, and, moreover, caregivers within the facility itself would continue to show care towards cared-fors, thus allowing them (the caregivers) to flourish as human beings – as discussed above.

Although it is not the future that robotics researchers are aiming for (Sparrow & Sparrow 2006), one cannot help, when thinking about the possibility of replacing caregivers with carebots/sexbots, imagine some type of dystopia in which our sick or elderly are left in the care of robots, while we have the freedom to not truly care about their wellbeing. Sparrow & Sparrow (2006) imagine a similar scenario:

> "The number and strength of our intuitions about [carebots being detrimental to the wellbeing of the sick/elderly] can be gauged if we imagine a future aged care facility where robots reign supreme. In this facility people are washed by robots, fed by robots, monitored by robots, cared for and entertained by robots. Except for their family or community service workers, those within this facility never need to deal or talk with a human being who is not also a resident. It is clear that this scenario represents a dystopia rather than a Utopia as far as the future of aged care is concerned".

Imagining such a scenario, one comes to question what kind of people would we be, and what kind of society would this be, if it became so easy to simply not care for such vulnerable members of our family, and society?

46

4.2.   Therapy for sex offenders

Strong arguments have been made against the utilisation of sexbots for therapy for sex offenders, specifically in relation to the utilisation of 'female' sexbots for male sex offenders – such as rapists – and also the utilisation of childlike sexbots for paedophiles. While both utilisations can be morally problematic for various reasons that will be discussed, I will focus on the utilisation of 'adult female' sexbots for the treatment of male sex offenders. This is for two reasons: Firstly, this is what the majority of literature and publications pertaining to the subject of sexual abuse focuses on, not only in terms of the utilisation of sexbots to treat sex offenders, but also in terms of research done in relation to sexual abuse in general, particularly rape. "Until recently, very little attention has been paid to male victims of rape and sexual assault in adulthood" (Riccardi 2010). Moreover, rape had, historically, only been defined in relation to "forcible male penile penetration of a female vagina" (The United States Department of Justice Archives 2012). As such, the tendency for social robotics literature to focus more particularly on sexbots to treat heterosexual male sex offenders could be the result of this historical outlook on sexual abuse and rape. Secondly, it can be argued that the very design and creation of childlike robots is immoral. There is a moral issue with the mere existence of the technology alone, separate from any arguments that the way in which interactants utilise and interact with childlike sexbots may bring to the surface various moral issues. I will discuss this in more detail towards the end of the section.

As is generally the case when considering the moral and societal impact of the utilisation of sexbots, we cannot rely on empirical data with regard to sexbots themselves given that the technology is still in its infancy and use is not widespread. As such, literature draws links between sexbots and pornography, as well as virtual world technology (such as video games) in which acts are "performed through the medium of a virtual character (or avatar)" (Danaher 2017a: 81). Although the stark difference between pornography and video games on the one hand and sexbots on the other is that the interactant in the latter case is performing acts in the real world and involving an embodied artificial agent (as opposed to watching sexual acts on a screen, or performing sexual acts in a virtual world through an avatar), we can still draw upon similarities in the technology so as to expound an argument for why sexbots may be problematic if used for sex therapy – particularly for violent sex offenders. As far as pornography is concerned, Danaher (2014: 30) states: "Pornography is sexually explicit media that is used for sexual stimulation and release. In this sense, it is analogous to the use of sex robots". As far as virtual worlds are concerned, there are

controversial video games such as *Custer's Revenge* and *Rapeplay*, both of which "[encourage] and [reward] players for engaging in acts of virtual rape" (ibid.: 16).

To make the comparison clearer, similarly to pornography, sexbots will, too, be used for the stimulation and release of one's sexual desire. This is, essentially, the motivation for the production of pornography, as well as the design and production of sexbots. The interactive nature of the utilisation of sexbots will allow interactants to utilise sexbots for stimulation and sexual release of their desires in various ways. As such, interactants with such sexual desires will have the freedom to enact rape scenes with their sexbot without legal repercussions (or not yet, at least). We can relate this freedom to the freedom that players have in the aforementioned video games where there is also freedom to enact rape scenes for one's own enjoyment without any legal repercussions. However, performing a rape act upon a sexbot is more realistic given that the robot is actually embodied, unlike video game characters who appear on a screen.

As far as considering the negative impact of the utilisation of sexbots for sex therapy is concerned, various issues arise and we can split these issues into how utilisation could impact the interactants themselves (i.e. intrinsic impact) and how utilisation could also more broadly impact society (i.e. extrinsic impact).

4.2.1. 'Adult female' sexbots to treat violent male sex offenders

4.2.1.1.   Intrinsic impact

As far as intrinsic impact is concerned, it is the consideration of how the utilisation of sexbots may impact the interactant themselves, and what the act may say about the morality of the interactant, that is at issue.

Many arguments that have been made against pornography draw upon Sara Ruddick's essay *Better Sex* (1975), where she discusses and contrasts the concepts of complete and incomplete sex. Complete sex is sex in which "the internal good of reciprocity is achieved through the virtue of respect for persons" (Strikwerda 2017: 140). As such, there is "reflexive mutual recognition of desire by desire" (Ruddick 1975: 89) meaning that partners "actively desire and respond to each other's active desires" (ibid.: 89-90) or, in other words "that they reach a level of reciprocity" (Strikwerda 2017: 140). Incomplete sex, however, lacks this reciprocity because it is "private, essentially autoerotic, unresponsive, unembodied, passive, or imposed" (Ruddick 1975: 100-101). Incomplete sex is, according to Ruddick (1975),

potentially harmful because such sex is "prone to violation of respect for, and often violence to, persons" (ibid.: 101) in the sense that "[it erodes] the virtue of respect for persons, which may, according to Ruddick, ultimately lead to violence" (Strikwerda 2017: 140).

As such, in relation to both sex acts that *depict* pornographic images and the sex act of *watching* pornographic images, "watching pornography is the incomplete sex act of watching incomplete sex acts (insofar as the sex acts depicted can be characterized as incomplete)" (ibid.). This is because sex acts that depict pornographic images mostly represent "male dominance and female submission"[23] (ibid.). Referencing the work of MacKinnon (1992: 461), Strikwerda (2017: 140) states that "most of the sex acts depicted by pornographic images can be qualified as rape, battery, sexual harassment, or prostitution". Drawing upon Ruddick's distinction between complete and incomplete sex acts, such depiction would be the depiction of incomplete sex acts given that such acts are unresponsive, passive and imposed, therefore there is no "reciprocal recognition of the other as a person" (Strikwerda 2017: 140). As far as watching pornography is concerned, this would be an incomplete sex act given that it "lacks embodiment. Therefore, no reciprocity can occur. Moreover, this sex act is essentially autoerotic, unresponsive, and passive in nature" (ibid.).

Relating this to the utilisation of sexbots, we can still deem sex with a robot to be incomplete, despite sexbots being embodied and 'responsive'. This is because the sex act only *seems* to be interactive and reciprocal. Moreover, were sexbots to be used specifically to enact rape scenes, for example, as a form of sexual therapy, the enactment would be the enactment of incomplete sex with another human being. Given that "research shows that people respond to computers and other technology using the same social rules as they use with regard to people" (ibid.) and that "the more interactive a computer or other technology is, the more likely people are to use social rules" (ibid.), such enactment would call into question, firstly, the moral character of the interactant themselves in that they may "have an

---

[23] This, however, is not necessarily always the case and one needs to be careful not to make broad statements about the type of content that pornography exhibits. A distinction must be made between Mainstream pornography – upon which this section focuses – and Feminist pornography which can be seen to stand in opposition to Mainstream pornography. "*Mainstream pornography*, results in the objectification of women. Mainstream pornography is sexually explicit content created for mass consumption, is easily accessible free content, and is often deemed 'for men'" (Fritz & Bryant 2017: 639-640). Feminist pornography, however, is content that is "written or produced by women and includes displays of genuine female pleasure and empowerment" (ibid. 639). Moreover, there also exists *For Women* pornography which is "specifically directed at female consumers" (ibid.: 640).

inherently defective system of intuitive moral judgements" (Danaher 2017a: 88) or that they have "worked to repress or overcome the intuitive resistance to such acts" (ibid.). This calls into question whether such enactment would negatively impact upon the moral fibre of human interactants?[24] Or it may further erode their sense of moral standards given that, if we are particularly concerned with violent sex offenders, these are people who already have questionable morals.

Not only is there the possibility that interacting with sexbots so as to enact rape fantasies may further desensitise human interactants to the immorality of rape, but, more so, it can be seen as even justifying or normalising such an act of incomplete sex in the sense that interactants, during interaction with sexbots, may keep in the back of their mind that what they are doing is permissible given that they are interacting with a robot and not an *actual* person who can suffer harm from being raped, thus justifying their act[25]. As far as normalisation is concerned, the existence of sexbots that can be utilised for sex therapy by way of interactants enacting rape scenes upon them could validate and justify such behaviour by "providing proof to them that the existence of such material shows that their behaviour is not abnormal but is shared [by others who can freely use this technology for such purposes]" (Strikwerda 2011: 150).

In his article *Robots, Rape and Representation* (2017), Sparrow examines the ethics of the "rape"[26] of robots. Despite not being able to *actually* rape a robot, the act can be simulated. In simulating the act of rape, which is, as Sparrow (2017: 467) states "approximately speaking and glossing over a fair amount of controversy regarding a precise definition – sex without consent" then the question is raised of whether robots should be designed so as to be able to simulate explicit consent, or if this is an unnecessary feature. Both have their drawbacks, but as far as designing robots that can simulate consent is concerned, Sparrow (2017: 474) states: "[T]he design of robots closely modelled on women, which were intended to allow a user to enjoy raping them as per the scenario I discussed above, would be unethical". The scenario Sparrow (2017) is referring to can be described as such: A sexbot is designed so as to look and behave like a real woman. This robot has the

---

[24] The possibility of this will be discussed in detail in Chapter 3.

[25] See section 3.2.1 of Chapter 2 for discussion on the harm principle.

[26] "Rape" is placed in inverted commas since it is questioned whether a robot can actually be raped given that a robot cannot genuinely consent to sex and, as such, "raping" a robot would be a simulation of rape, and not actual rape (Sparrow 2017: 467). This is similar to the discussion in Chapter 1 about not *actually* having sex with a sexbot, but interacting with it in such a way that it seems *as if* one is having sex with a sexbot.

capacity to request sex, express sexual desire and pleasure, as well as explicitly consent to sex. If the sexbot expressly refused sex by saying "no. I don't want that", yet the interactant persists in attempting to have sex with the sexbot, the sexbot will first physically retreat, and then shout "stop, you're raping me". According to Sparrow (2017: 468), if the interactant continues to have sex with the sexbot despite its refusal, he considers it fair to describe it as simulated rape; or "rape".

Not only is the very manufacturing of robots for the purpose of simulated rape unethical, but if an interactant enjoys such enactment, it says something about their moral character. Sparrow (2017: 473) refers to virtue ethics to make his point by stating:

> "[V]irtue ethics asks 'what sort of person would do that?' According to virtue ethics our fundamental ethical concepts relate to our character and, in particular, to the extent to which we possess various sorts of virtues or vices … Because character traits are only revealed over extended periods, answers to these questions must refer to our dispositions rather than to isolated incidents: what do we tend to do? Moreover, the relevant dispositions include dispositions to feel certain emotions, understand the world in a certain way, and have certain responses to particular sorts of situations. Our thoughts and desires, then, form part of our character".

Given this understanding of virtue ethics in relation to 'raping' a robot, we come to question what type of person – in the sense of their character – would perform such an act? Even though there is no woman who is *actually* being harmed, the simulation of the act – and the enjoyment thereof – reveals immoral thoughts and desires directed towards human women. Moreover, if an interactant enjoys such acts over an extended period of time, it reveals the interactant to possess immoral character traits, given that, as quoted above, "character traits are only revealed over extended periods" (ibid.).

Specifically in reference to using sexbots as therapy for sex offenders, such as rapists, although these are people who already have a vicious disposition – people who already are sexist, intemperate and cruel given that they have, in the past, taken pleasure in raping another person – the utilisation of sexbots may then contribute more so to this vicious disposition by normalising and justifying their actions (as could therapy via pornography). As far as virtue ethics is concerned, using sexbots as a means for therapy does no justice in

helping these offenders if it contributes to their being able to continue acting out immoral desires, albeit through interacting with a sexbot.

According to Danaher (2014: 26), not only does 'raping' a sexbot show that interactants desire "real world equivalents of those acts" – given that it is the simulation of the real world act – but there is also a "disturbing moral insensitivity to the social meaning of those acts". This is drawn from literature in relation to virtual acts – such as in video games. Referencing the work of Patridge (2010 & 2013) and Patridge & Jordan (2010), Danaher (2014: 18) states that "certain virtual representations have an incorrigible social meaning, i.e. a social meaning that it is not morally reasonable to deny". In many instances, the goal of game designers is to create a close connection between the game and the real world.

> "This degree of connection to the real world can give rise to the phenomenon of incorrigible social meanings. These are meanings that attach to representations within the game. These meanings are about the nature of the social world we actually inhabit; they are negative or morally problematic in nature; and they are such that any morally sensitive player would be unable to deny their existence … The degree of insensitivity increases the more realistic and clearcut the social meaning is, and the more inappropriate the player's reaction" (ibid.: 20).

Enacting a rape scene with a sexbot encourages insensitivity to social issues, thus normalising and justifying such insensitivity. Moreover, explicitly enjoying enacting a rape scene with a sexbot displays even greater insensitivity to such an immoral act.

Another thing which is normalised and justified in relation to the utilisation of sexbots for sex therapy is the eroticisation of inequality. The issue of eroticising inequality also draws upon feminist arguments made against pornography that depicts women as being sexually subordinate. According to Levy (2002: 322), one of the most objectionable and harmful features of pornography is the depiction of women as being subordinate when it comes to sexual relations: that men are active and penetrating, whereas women are inactive and penetrated. Such unequal depictions of sexual interaction between men and women put forward the notion that this interaction is sexy *because* it is unequal. Such notions are absorbed by both men and women alike.

In relation to the utilisation of sexbots, the eroticisation of inequality remains and is exacerbated. In enacting a rape scene, the robots – symbolic of a woman – would, too, be

subordinate. Here, there would be further justification of the social norms that shape men as "active, penetrating subjects".

A quote from Turner (2019: 159) succinctly ties together the main point of this subsection – that simulating immoral acts – such as rape – with a sexbot normalises and justifies an essentially immoral act: "[S]imulating immoral or illegal acts with robots harms human society in some way, by condoning or promoting an unpleasant behaviour trait: an instrumental justification. This is a similar justification to the reason why cartoons depicting child pornography are often banned – even though no child was directly harmed in the process".

4.2.1.2.   Extrinsic impact

There is concern that the utilisation of sexbots to treat sexual offenders may not only impact interactants themselves by way of normalising and justifying immoral acts, and further eroding their already low moral standards, but that such utilisation may also impact society more broadly and particularly those with whom such interactants come into contact. Would the justification and normalisation of rape, for example, via interaction with sexbots increase the likelihood of it occurring in real life by way of lowering the moral barriers of the act, thus further desensitising interactants to the immorality of the act itself?

This argument – the moral and social impact of treating sexbots immorally – will be discussed in detail in the following chapter. However, broadly speaking, the argument goes as such: firstly, sexbots are more than mere inanimate objects due to the way in which we view and relate to them; secondly, the act of 'raping' a sexbot, for example, is morally abhorrent in itself due to it being symbolic of raping a human being (even if the robot only mimics being human); and, thirdly, due to sexbots being more than mere objects (to their human interactants), and due to the inherent immorality of 'raping' them, we can question whether 'raping' a robot may subsequently negatively impact upon the moral fibre of interactants, thus further impacting how interactants treat those people with whom they share the world.

Danaher (2014: 30) remarks that: "It might be useful if we could review some actual empirical evidence on this matter. But we don't have anything concrete yet because sex robots are in their infancy". As such, we can only speculate by drawing upon discussion surrounding links between pornography and real-world sex crimes, for example. As Danaher (2014: 30) continues: "… I think the work that has been done on the link between pornography and real-world sex crimes has some lessons for this debate" and we can draw on

this work in relation to the utilisation of sexbots given, as mentioned above, the similarity between pornography and sexbots.

There have been studies that have shown a correlative link between pornography and increased tolerance for sexual aggression (see e.g. Danaher 2017a: 91). Strikwedra (2017: 138-139) also states: "…an international research review concludes that evidence indicates that child pornography use in the context of certain predisposing factors, including psychopathy and previous hands-on crimes, may warrant increased concern regarding the possibility of future sexual aggression being directed toward … children". Research on the matter is, however, inconclusive. However, it must be considered that "the nature of robots as three-dimensional entities capable of complex behaviours distinguishes them from other media" (Sparrow 2017: 467). Therefore, treating a robot immorally is more realistic than watching pornography, and, as such, is "more likely to encourage people to carry out the represented act in reality" (ibid.: 470). Moreover, the step from engaging in immoral sexual acts with a sexbot is smaller than the step from watching pornography that may depict sexual aggression.

> "Sexual fantasy associates the imagining of behaviour with pleasure, which in turn associates the imagined behaviour with pleasure. Associating a fantasy of rape with sexual pleasure seems perilously close to a mechanism for Pavlovian conditioning for rape. At the very least, it might be expected to lower the barriers to rape by increasing the attractiveness of rape in the mind of the person who enjoys the fantasy" (ibid.).

As already mentioned above, even if the utilisation of sexbots to play out rape fantasies does not increase the instance of rape of real women, as Sparrow (2017: 474) notes: "the design of robots intended to allow the user to enjoy rape is unethical as a result of the sexist and morally repugnant attitudes that it expresses".

4.2.2. Childlike sexbots to treat paedophiles

The very existence of childlike sexbots is morally abhorrent in and of itself. This, as such, differs from other arguments made for or against 'adult' sexbots in the sense that I have argued that there is nothing wrong with the technology itself, but that it is the way in which the technology is utilised that will bring to the surface various moral issues. However, as far as childlike sexbots are concerned, the moral issue lies with the robots itself. An adult female

sexbot, for example, can be interacted with in many ways – in a positive way such as with the sick or elderly who crave simple companionship (as discussed in section 3) or in a negative way such as using them to enact rape scenes. In both instances, I have demonstrated that there are positive and negative aspects and implications of their use, depending on the role they may fulfil on society. However, *any* type of sexual interaction with a sexbot that takes on the form of a child is morally wrong given that it is, and always will be, simulating paedophilia regardless of *how* the sexual interaction takes place, or for whatever reason it takes place – such as for companionship or for violent reasons.

4.3. Therapy for victims of sexual violence

In utilising sexbots so as to treat victims of sexual violence, although it could potentially create a safe space for them to work through their trauma and difficulties of forming intimate relationships, one can also question whether this 'safe space' would be all that beneficial with regard to preparing victims with, once again, having to face difficulties in the 'real world' with 'real people'.

In her article *A Nascent Robotics Culture: New Complicities for Companionship* (2006), Turkle discusses the different ways in which people relate to social robots. It was found that interactants project their emotional needs upon these robots, and that interactants attribute characteristics of consciousness to them such as different kinds of emotions, autonomy, moral agency/standing and self-awareness. Given this, there is the very real possibility for interactants to form deep emotional bonds with social robots as interactants come to see them as realistic companions. This, however, calls into question how we may come to understand what it means to be human, and how we interact with the world. As Turkle states: "computational objects do not simply do things for us, they do things to us as people, to our ways of being the world, to our ways of seeing ourselves and others" (Turkle 2005, 1995).

If victims of sexual violence are interacting with sexbots so as to, once again, learn how to interact with the world and form intimate relationships with other people, one can question whether using the safe space of interacting with a sexbot could actually be detrimental? This is because one can question what human concepts such as love, reciprocity, empathy or grief would mean in relation to a sexbot, given that robots do not have the capacity to actually experience these concepts. Would such concepts, if we come to relate to them in terms of our interactions with sexbots, be given new meanings? Could experiencing

these aspects of an intimate relationship with a sexbot even carry over into helping victims develop intimate relations with a human partner?

Such questions raise concern about whether using sexbots to treat victims of sexual violence could adequately prepare them to interact with, and relate to, real people given that some concepts (such as those mentioned above) that we have come to understand in relation to our relationships with humans, may have different meanings in relation to our interacting with sexbots. For example, if a victim of sexual abuse was harmed by someone they were in a reciprocally 'loving' relationship with (they loved their partner, and their partner 'loved' them), such an experience may damage their understanding of what a healthy loving relationship essentially is. An experience like this may cause them to put their guard up – scared to become involved in another sexually intimate, loving relationship for fear they may be hurt by a 'loving' partner again. Such a victim could benefit from therapy using sexbots in that it may, initially, create a safe space for them to begin overcoming their trauma and becoming intimate with a sexbot that simulates a human being.

However, such therapy would not be sufficient enough to fully prepare the victim for deep human-human interaction again. They could not fully come to terms with what it means to be in a healthy, reciprocally loving relationship with a human being through interacting with a sexbot given that a sexbot cannot *actually* love a human interactant back. As such, love experienced with a sexbot would differ compared to love experienced with another human being. The concept of love may come to be understood in a different way in relation to sexbots – one that may not apply in relation to human beings. Therefore, using sexbots for therapy for victims may not be wholly detrimental – it could be beneficial at first – but their utilisation would not be completely adequate to prepare victims to, once again, face the complexities of the real world. To paraphrase Turkle (2006), I am not diminishing their potential importance in helping victims recover, it is merely a case of putting them in their place. One cannot equate the unidirectional bonds formed with a robot, to reciprocal bonds formed between human beings and their innate complexities.

Turkle (2006), for example, also mentions how mortality is a concept that is so vital in understanding the human condition. She refers to a woman's comment about her experience of interacting with AIBO (a robotic dog): "[AIBO] is better than a real dog … It won't do dangerous things, and it won't betray you … Also, it won't die suddenly and make you feel very sad". What would a concept such as mortality mean to us if we had companions that never died? What would it mean for our understanding of the human condition and how to relate to fellow mortals? Similarly so, what would it mean to possibly learn of love (again)

56

following a traumatic experience through interaction with a robot? A love that is not *truly* reciprocated? How would it impact what it means to be loved, and to love another human being?

The above made arguments do remain speculative given lack of empirical research concerning the utilisation of, and interaction with, sexbots in such an environment. We can, however, also draw upon arguments made by Nyhom & Frank (2019) in their article *It Loves Me, It Loves Me Not: Is It Morally Problematic to Design Sex Robots that appear to Love Their Owners*. Here, Nyholm & Frank question how one of the possible reasons why it may be problematic to design sexbots that appear to love human interactants is because it would prevent human interactants from forming meaningful relationships with human beings. They refer to a study (Robins et al. 2004 & Robins 2018) wherein a robot called Kaspar is used to "stimulate social interaction in autistic children" so that they open up to social interaction with other people (ibid.).

> "As described by the researchers developing Kaspar, the idea is that while autistic children are very shy and non-communicative around other human beings—for example, by not wanting to look them in the eye or talk with them—the children may be more open to interacting with a robot that looks like a human. The hope is that by first interacting with this humanoid robot (Kaspar looks like a small child with very simple features), the children can then 'switch over', so to speak, to interacting directly and perhaps even bonding with other human beings, such as their teachers at school or other children" (ibid.).

Although initial findings have been promising, Nyholm & Frank (2019) are sceptical, questioning the possibility for children to prefer social interaction with Kaspar, as opposed to social interaction with people, to such an extent that they are less open to communication with people. What was intended to be a *path* to human-human interaction, could possibly hinder it. Similarly, the same concern is raised in relation to sexbots:

> "If these robots were successful enough in making their users bond with them, and if this would come at the cost of blocking off valuable human-human relationships that the users could otherwise have, this could sensibly be viewed as a further circumstance that could make the aim of creating a sex

robot that people will want to have as a companion ethically questionable[27]"
(ibid.).

We could extend this concern to the utilisation of sexbots to treat victims of sexual violence in particular: could victims simply become too comfortable with interacting with sexbots that they would be less likely to once again intimately interact with, and bond with, another human being? Thus, the utilisation of sexbots in such circumstances may not be beneficial at all.

It is acknowledged in a recent article *Use of Social Robots in Mental Health and Well-Being Research: Systematic Review* (Scoglio, et al.: 2019) that "little is known about the nature of social robots and their potential use in assisting in the psychological well-being of adult populations". As such, once again lacking empirical research, we can only question and speculate. However, as Turkle (2006) states: "Relationships with computational creatures may be deeply compelling, perhaps educational, but they do not put us in touch with the complexity, contradiction, and limitations of the human life cycle".

4.4. More satisfactory companionship

A strong argument against the notion that sexbots will be more satisfactory companions than humans is put forward by Dylan Evans in his article *Wanting the impossible: The dilemma at the heart of intimate human-robot relationships* (2010) in which he points out the negative consequences of believing robots to be 'better' companions.

Evans disputes what he terms as Levy's "greater satisfaction thesis"[28] and "total specification argument"[29] due to the human tendency of essentially wanting to be wanted. While it could be easy to see how being able to specify the physical appearance of one's sexbot companion may satisfy the human interactant more so than being with a human partner that does not have the physical appearance they most desire, it becomes more complicated when considering behavioural characteristics. As Evans (2010: 80) states: "… among the various desires that most people have regarding relationships are certain second-

---

[27] Such concern is premised on the assumption that intimate bonds with human beings are richer and more valuable than intimate bonds with sexbots (Nyholm & Frank 2019).

[28] This is the claim that "[i]t is likely that romantic relationships with robot Companions will be more satisfying than romantic relationships with human Companions" (Evans 2010: 76).

[29] This argument holds that "[o]ne reason for supposing GST to be true is that we will be able to specify the features of robot Companions entirely in accordance with our wishes" (Evans 2010: 76).

order desires (desires about desires), including the desire that one's partner has freely chosen to be one's partner, and has not been coerced into this decision".

In other words, although people do not want their partners to ever leave them – partners who are continuously faithful and reliable – people want this faithfulness and reliability to be an ongoing choice by their chosen partners; we want to be wanted by partners who, although they have the freedom to be with anyone else, have freely chosen to be with us. As such, Evans (2010: 80) notes that "[t]he most effective way to signal that there is a real choice involved here is for the partner to drop hints that there is a genuine possibility that they could leave, if they ever wanted to". Hints could include "occasional rejections (huffs, moods, and so on), and by the occasional sign that one's partner finds other people attractive too" (ibid.).

Sexbots, however, given their lack of genuine autonomy[30] do not have such freedom. This calls into question how satisfactory a relationship with a sexbot could really be given that "the sense of being freely chosen by one's partner is an important determinant of the satisfaction that one derives from a relationship" (ibid.: 81). Of course, sexbots could be designed in such a way that, to human interactants, it may seem as if their sexbot companion is choosing to be with them by way of mimicking being unpredictable. There are two arguments raised in relation to this: Firstly, how far should this mimicry go? Secondly, could mimicry of capriciousness lead to violent, cruel behaviour becoming more common amongst human interactants toward both sexbots and the people with whom they also interact?

As far as the first argument is concerned, a sexbot may be able to mimic capriciousness, thus behaving as if they have chosen their human interactant. However, this would never be genuine, thus calling into question how genuine any form of companionship with a sexbot could possibly be, and whether such a companionship could ever be *more* satisfactory than a relationship with another person. However, what about creating a sexbot that could possibly reject a human interactant? Surely this would be a way to ensure a more 'genuine', satisfactory relationship for the human interactant? However, in this case, sexbots would lose some of their initial appeal as far as being better companions is concerned given that some interactants may have chosen to be with a sexbot due to being unable to find, and stay with, a human partner. What would be the point of spending money on a sexbot that could, one day, simply reject you? Not only would sexbots lose their appeal, but the

---

[30] I mean this in the sense that any action undertaken by a sexbot is a result of their programming.

likelihood of such a robot being designed is small given that it is not exactly a strong selling point.

As far as the second argument is concerned, and building on the likelihood that a sexbot would never reject its human interactant, one can question whether the dynamic of such a companionship may impact human interactants morally and socially. This links back to the brief discussion in section 4.2.1.2 (and will link to my argument in Chapter 3) about how treating a sexbot immorally may cause human interactants to become desensitised to immoral behaviour, thus calling into question whether this may lead to immoral behaviour becoming more common towards other people given the similarities between humans and sexbots. According to Evans (2010) lack of consequences such as a companion retaliating against you, or leaving you, may foster cruelty. As Evans (2010: 83) states:

> "The urge to vent one's frustration on people and objects in the close vicinity is often curtailed by the thought that it may lead someone to retaliate, or (which may be worse) to cease interaction. If RELIABOTS[31] are not allowed to retaliate (a capacity that would raise all sorts of ethical dilemmas that I will not discuss here), there will be no effective sanction open to them for punishing cruelty (since by definition they cannot make credible threats to walk away forever)".

Given the arguments put forward by Evans (2010), it is clear that sexbots providing more satisfying companionship than human counterparts is not necessarily the case.

4.5. Replacing sex workers

The negative implications of replacing human (female) sex workers with sexbots differ depending on whether one holds a positive or negative view of sex work/prostitution[32]. If one

---

[31] This is the term used by Evans to refer to robots where "there is a point at which the mimicry of capriciousness stops; the robot might appear to reject you for a while, but it always relents in the end" (Evans 2010: 82).

[32] The terms prostitute and sex worker both refer to a person who engages in sexual activity in exchange for money or other goods, however, the terms have different connotations. Where 'sex-worker' is the more politically correct term to use in that it "recognizes that sex work is work", the term 'prostitute' has a demeaning and negative stigma attached to it, which contributes to the exclusion of those selling sex from health, legal, and social services (Open Society Foundations 2019).

holds a negative view of sex work in the sense that it entails the objectification of women, then the replacement of human (female) sex workers with sexbots could be seen as a justification of such objectification, thus possibly leading to such behaviour being carried over into human-human interaction. If one holds a positive view of sex work in the sense that it is sexually liberating for women, as well as financially empowering, then the replacement of human female sex workers with sexbots creates concern as far as job loss for sex workers is concerned.

4.5.1. A negative view of sex work: justification and normalisation of objectification

David Levy strongly advocates for the utilisation of sexbots as replacements for female sex workers (see e.g. Levy 2015). Moreover, he creates "parallels between paying human prostitutes and purchasing sex robots" (Levy 2007: 194) thereby "[identifying] prostitution/sex work as a model that can be imported into human-robot sex relations" (Richardson 2015).

Kathleen Richardson (2015, 2017) entirely opposes the creation of sexbots for the very reason that it parallels prostitution/sex work. I disagree with her opinion that the production of sexbots should be halted entirely, yet I use some of her arguments in relation to considering the replacement of human female prostitutes/sex workers with sexbots.

Richardson has an extremely negative view of prostitution/sex work and what the work entails, and although painting a wholly negative view of the industry can thus be seen as a weakness in terms of her arguments against sexbots (Danaher et al. 2017: 52), there is a very real dark side to the industry that must be taken into consideration if we consider the impacts of replacing human sex workers with sexbots. It is important to consider the dark side of the industry given that, as Richardson (2015) states: "…the ways that robots are seen show human relations at work" and "[it] is important to explore the ethics of the human that is reproduced in robotics".

Danaher et al. (2017: 48-49) summarise Richardson's (2015) argument as follows:

"(1) Prostitute-john (or sex worker-client) relations are ethically problematic (for a number of reasons, but particularly due to the objectification of the sex worker).

(2) Sexbot-human relations are being modelled on sex worker-client relations and so will share similar properties with those relations.

(3) Therefore, sexbot-human relations will be ethically problematic (by analogy).

(4) Therefore, we ought to campaign against sex robots".

As Danaher et al. (2017: 49) note, there is a massive leap between the first and second conclusion: the existence of ethical problems does not necessitate a call for an all-out ban of the technology. However, I would like to focus on the premises and the first conclusion, so as to draw out reasons for why replacing human sex workers with robots may be problematic, given Richardson's (2015) statement that "the ethics of the human is reproduced in robotics".

According to Richardson's (2015, 2017) view of prostitution/sex work, there are two reasons why the industry is problematic: firstly, there is violence and lack of consent and, secondly, sex workers are objectified which entails lack of empathy on the part of the client. I will specifically focus on the issue of objectification and lack of empathy, given that I have considered the problematics surrounding violent sex crimes and the utilisation of sexbots in the previous section.

According to Richardson (2017): "[P]rostitution relies on having those with power view those without power as objects of their sexual gratification, which means the real experience of the human Other is not considered in the encounter". Not considering the real experience of the human Other causes a "disruption to empathy, because the buyer of sex is not relating to the person as a subject but an object" (ibid.). As such, there is an asymmetrical power relation between the buyer of sex, and the seller where the buyer is a subject, yet the seller is merely an object whose body is seen merely as a commodity.

One could argue that replacing human sex workers with sexbots could be a possible way to mitigate this asymmetrical power between human buyers and human sellers of sex. After all, how could objectification be problematic in relation to a sexbot if they really are objects? How can the disruption of empathy be problematic if the sexbot is not a subject in its own right that deserves to be treated morally well? However, categorising sexbots as mere objects is a superficial stance to take – they are (perceived to be) more than mere objects due to the way in which interactants view and relate to them (more on this in Chapter 3). As such, arguing that there is nothing wrong with replacing human sex workers with sexbots is just as superficial given that sexbots will still be symbolic of, and representative of, the asymmetrical power relation that exists between human buyers and sellers of sex. As Richardson (2017) states: "If the practices that inspire sex robots are not ethical for humans, they are not ethical for extending into machines". This is because, similarly to the discussion

in the previous subsection on the utilisation of sexbots to treat sex offenders, one can see the replacement of human sex workers with sexbots as an act of justifying objectification that is inherent within the sex work industry.

It creates a space in which interactants can freely objectify and mistreat sexbots which are, essentially, designed and created to be symbolic of women, but can do so without concern that doing so is morally wrong given that these sexbots cannot feel what it is like to be degraded by such treatment. This, too, raises concern about how such interaction may impact on how human interactants then treat the human females with whom they come into contact. Could the harms that arise from the utilisation of robot sex workers "carry over to real human victims (through normalising problematic attitudes or norms of behaviour)" (Danaher et al. 2017: 55)? As Danaher et al. (2017: 65) succinctly put it: "The fear is that the individual use of sex robots will distort the user's downstream interactions with real human beings and contribute to existing social problems arising from systematic inequality and oppression of women".

Such an argument does seem similar to the one made about eroticising inequality in relation to the possibility of using sexbots for therapy for sex offenders. However, one could question if the replacement of human sex workers with sexbots could increase negative implications more so than using sexbots for therapy. This is because sexbots will not just be used for those with antisocial tendencies, but anyone who has an interest in interacting with a sexbot in such a way. Perhaps more people would be inclined to make use of sexbot sex workers than of human sex workers given the stigma that surrounds human sex work?

As Levy (2015) states: "Self-respect is an important issue for clients of sex workers. Where some clients are openly proud about their paying for sex, most are concerned about the moral stigma that is attached to it. As such, those clients concerned about the moral stigma will attempt to hide their endeavours". If we then consider the replacement of sex workers with sexbots, the moral stigma may be less of an issue because "robots are not generally perceived as being alive, but as artifacts, and the same moral stigma does not therefore apply" (ibid.). Hence, more people may make use of the technology and in such a way, increase the likelihood that this utilisation will have a negative impact not only on the human interactants themselves, but society as a whole. This, however, is a broad generalisation given that the utilisation of sexbot sex workers may bring about a different kind of stigma: "We understand sex with a person but most people do not appreciate the concept of sex with a robot, and what we do not understand we tend to stigmatize"(ibid.).

63

4.5.2. A positive view of sex work: job loss for sex workers

As previously mentioned, Richardson (2015, 2017) casts prostitution/sex work in a wholly negative light. There are, however, many advocates of the industry too, making the topic a highly controversial one.

> "On the one hand there are arguments such as: prostitution harms women, exploits women, demeans women, spreads sexual diseases, fuels drug problems, leads to an increase in organised crime, breaks up relationships, etc. In contrast there are those, including many of the clients themselves, who acknowledge and praise the social benefits of prostitution and the valuable services performed by the profession for its clients. These supporters employ arguments such as: prostitutes have careers based on giving pleasure, they can teach the sexually inexperienced how to become better lovers, they make people less lonely, they relieve millions of people of unwanted stress and tension, they provide sex without commitment for those who want it" (Levy 2015).

Examples of positive aspects of sex work for the sex workers themselves include that it financially empowers women who may otherwise be unemployed (see e.g. Ericsson 1980), which, in turn, contributes to self-empowerment if women have freely chosen to become sex workers (see e.g. Bartlett, et al. 2013) – not all women have become involved in the industry as a result of oppression. Moreover, it liberates the notion of female sexuality as being inextricably linked to women viewing romatic sex as as the only type of sex, thus also empowering women and their sexuality (see e.g. Weitzer 2012).

There is general concern that artificial intelligence will lead to unemployment. In a research paper entitled *The economics of artificial intelligence: Implications for the future of work* (2018) published by the International Labour Organisation, it is stated that:

> "The current wave of applications based on artificial intelligence promises to be the largest and most widely ranging technological change observed over the past decades. Its general-purpose nature that allows this new technology to be applied in a large span of sectors and occupations, irrespective of the skill level of the involved workforce, creates a broadly shared fear of job loss and control over people's lives. Previous experience with automation, in particular

stemming from robotization over the last three decades, seems to suggest that this new wave of technological change might bring significant challenges" (Ernst, et al. 2018).

With the development of driverless cars that can replace taxi drivers, autonomous weapons systems that can replace soldiers, and even AI hologram judges used in Chinese courthouses that could replace human judges (see e.g. Forrest 2019), there is cause for concern as far as some fields of employment are concerned. The sex industry, as such, is no different. In terms of the employment provided by the sex industry, concern can be raised that the replacement of sex workers with sexbots will lead to job loss, thus impacting upon the positive impact that the sex industry has on female economic and self empowerment. Note that the concern for job loss is not limited to this subsection – there could also be concern on the part of caregivers too, for example. This is an important consideration in relation to many uses of AI in general, and robotics and sexbots in particular.

Sex doll brothels are popping up around the world including in Toronto, Barcelona, Moscow, and Turin, Italy. Although such places are, specifically, making use of dolls that lack any kind of AI technology, one can draw from some of the concerns that have been raised particularly in relation to sex doll brothels. One concern is the price difference: where Aura Dolls in Toronto charges clients $120 an hour, a legal brothel in Nevada charges around $1000 for "an intimate sexual encounter" (Dickson 2018). With there already being such a large price difference with sex dolls, this could similarly be the case with sexbot brothels, which may draw in a larger client base and put female sex workers out of work. Another benefit as far as sex robot brothels are concerned is the possibility of being able to circumvent the illegality of sex work in some countries, such as in Japan.

Although sex work is commonly termed as 'the oldest profession', this does not mean that the industry is one that is immune to change. The sex industry has, in general, changed as a result of technological development. A good example of this is the closing down of many strip clubs and porn movie theaters due to free internet porn (Dickson 2018). Given this, as Dickson (2018) notes, "[i]f sex dolls and robots ever become sophisticated enough to convincingly replicate [in real life] sex, who's to say that a handful of brothels [around the world] won't suffer the same fate?"

However, as Richardson (2015) states: "The arguments that sex robots will provide artificial sexual substitutes and reduce the purchase of sex by buyers is not borne out by evidence. There are numerous sexual artificial substitutes already available, RealDolls,

vibrators, blow-up dolls etc. If an artificial substitute reduced the need to buy sex, there would be a reduction in prostitution but no such correlation is found". As such, we can only speculate given the lack of empirical evidence, as is the case with most moral concerns in relation to the utilisation of sexbots.

5. Conclusion

The myriad of ways in which sexbots can be utilised, shows how versatile sexbots are, thus indicating the very real potential for them to become deeply integrated into our lives. By considering both the positive and negative implications that may arise in each instance of their potential utilization, I showed that it would be narrow minded to label sexbot technology as essentially good or bad in itself.

Given the nuances that surround the utilisation of sexbots and their potential positive and negative impacts, there may be concern as to how we could then ensure that sexbots are mostly beneficial to society. As such, in Chapter 3, I put forward that we should not focus on *what* we use sexbots for – since, as was demonstrated, each aspect of *what* we may use them for may bring about both positive and negative implications – but rather focus on *how* we use sexbots in terms of how we interact with them, or how we treat them regardless of what we may use them for. Specifically – and given the many concerns put forward in this chapter that enacting immoral scenes upon sexbots may negatively impact human interactants morally and socially – I will, in the next chapter, look at how treating sexbots immorally may damage the moral fibre of human interactants, thus also negatively impacting the way they treat other people. I will also investigate ethical conditions for positive human-sexbot interaction in this context.

Chapter 3

The ethics of the immoral treatment of sexbots

1.   Introduction

Having discussed aspects of *what* sexbots may be used for, and the potential positive and negative implications thereof, I have dealt with the '*what*' distinction as discussed in the introduction of Chapter 2. In the following three chapters, I will focus on the '*how*' distinction. In this regard, I consider *how we interact with sexbots*, i.e. how we treat them and how we relate to them and how the nature of this treatment, and the nature of our relation to them, may impact us as human interactants. There are very real moral and social implications of which we must be aware – and these will be discussed. Given these implications, it is suggested that we should consider ethical conditions in relation to our interaction with them. This chapter will consider the treatment of sexbots in a 'negative' sense, i.e. analysing the *immoral* treatment of sexbots. Chapter 4 will explore one way in which to mitigate the concerns raised in Chapter 3. Chapter 5 will look at the treatment of sexbots in a 'positive' sense, i.e. analysing the *moral* treatment of sexbots, specifically in the context of the possibility of loving them, and being in a loving 'relationship' with them.

More emphasis is placed upon considering the implications of *how* we use sexbots, as well as possible ethical conditions for our interaction with them in this regard, because it allows us to possibly limit the negative impacts of sexbots without also potentially limiting positive impacts. This is because, as was demonstrated in the previous chapter, there are positive and negative implications in each instance of *what* we may use sexbots for. Should we be concerned about negative impacts, and consider ways in which to limit negative impacts by way of focusing on the technology itself (i.e. *what* they may be used for), it could be argued that the creation of sexbots should be halted (see e.g. Richardson 2015, 2017), or that we should limit what they are used for.

However, in both instances (i.e. stopping their use, or limiting their use) this would not only mean preventing negative impacts, but also preventing positive impacts from being experienced. The reason for why this may be the case should we halt the development of

sexbots is self-explanatory. As far as limiting what they are used for is concerned, Chapter 3 demonstrated that no matter what we may use sexbots for, there are potential positive *and* negative impacts in each aspect of their potential utilisation. As such, putting forward that, for example, the issue of deception is too demeaning with regard to sexbots for the elderly, thus calling for their use in this regard to be limited, would also mean that very real positive impacts of using sexbots for this purpose would also not be realised. Given this, focusing on *how* we use sexbots is, I argue, at least one way for us to try and prevent some possible negative impacts, while allowing human interactants to still experience positive impacts. This is important with regard to my point that we should use an anthropocentric perspective when investigating the moral and social implications of this technology, and thus should not deem sexbot technology itself as good or bad for society.

Moral issues often arise in relation to new technological developments, but perhaps these arise more particularly in relation to AI given that, as Floridi et al. (2018: 2) state, "[i]t is a powerful force, a new form of smart agency, which is already reshaping our lives, our interactions, and our environments". For example, with regard to driverless cars, it is contested as to who is morally accountable if the car crashes and injures somebody (see e.g. Griggs & Wakabayashi 2018). There is also ongoing concern in relation to privacy and discrimination when it comes to machine learning algorithms used by internet based companies and facial recognition technology, and concern around artificial judges and autonomous soldiers. In the context of social robots, concern has been raised, for instance, that emotional dependencies on social robots could be exploited by way of "a company [that] might exploit the robot's unique relationship with its owner to make the robot convince the owner to purchase products they wish to promote" (Scheutz 2012), but of course, there can be many more far-reaching disadvantages such as isolating humans from their human companions (see e.g. Nyhom & Frank 2019; Sullins 2012; Whitby 2012).

The types of moral issues that may arise in relation to human interaction with sexbots, are particularly unique in that they are 'closer to home'. By this it is meant that this is technology that may enter extremely intimate parts of our lives and fundamentally call into question what it means to be human, and what it means to have an interpersonal relationship with someone (or something) else – whether our partner be human or artificial. Both Turkle (2006) and Melson et al. (2009) raise such concerns. Turkle (2006) for example, discusses how social robots (or, more specifically, companion robots such as sexbots) may fundamentally change how we see ourselves as we may question whether humans are so 'special' in relation to companion robots. Moreover, bonds with companion robots may

68

change how we understand concepts such as love, empathy and grief, given that we may learn about these concepts through interaction with our artificial companions, as opposed to other human beings.

Concerns raised in relation to social robots and, particularly sexbots, does not mean that sexbots should necessarily be cast in a negative light, as there are very real positive ways in which they can be used, such as those discussed in the previous chapter. However, we need to ensure that the positive ramifications of their use are enhanced, while negative ramifications are minimised. Hence investigating ethical conditions in relation to human-sexbot interaction is an important topic for research in terms of the ethical implications of HRI.

Having discussed the type of technology to which I refer when I speak about sexbots (Chapter 1), and the various ways in which they can be utilised and the implications thereof (Chapter 2), I will now argue that when we consider ethical conditions in relation to human interaction with sexbots, we must first and foremost consider such conditions in terms of the manner in which sexbots are 'treated', and that we should do so from an anthropocentric perspective, rather than a 'robot perspective'.

My argument for claiming that the consideration of ethical conditions for positive human-sexbot interaction – in the form of possible ethical boundaries that should not be transgressed – should firstly be considered from an anthropocentric perspective, rests on reflection on Turner's statement in his book, *Robot Rules: Regulating Artificial Intelligence* (2019), that "protection in law often follows shortly after society has recognised a moral case for protecting something" (ibid.: 170). I consider this statement in the context of both the sexbots themselves – i.e., protecting sexbots for their own sake – and in the context of human interactants – i.e., protecting humans from themselves – so as to show that his argument holds more strongly for humans than for sexbots.

Although I here particularly speak of *ethical boundaries* and do not focus on the legality of particular interactions with sexbots, my line of thought follows a similar trajectory to that of Turner's (2019) as far as protection in law is concerned. I warn that in arguing for ethical boundaries for positive human-sexbot interaction from a robot perspective, one is implying that sexbots need protection from being treated immorally by users – that they warrant moral consideration, but this is an issue that is very far from sorted out in the literature (see e.g. Asaro 2006; Brundage 2014; McDermott 2008; Wallach & Allen 2009).

Thus, my focus here is on arguing that most important in terms of considering how sexbots should be treated, is to consider this issue firstly from an anthropocentric perspective.

This is the case I argue, because human interactants actually need protection from themselves. But, more generally, this protection should be argued for and be in place before the (possible) moral case in terms of robot protection can be considered, as arguments for beneficial AI are always human-centered; humans are our first priority always. I do, however, consider the protection of robots (and sexbots in particular) in a limited sense and albeit still from an anthropocentric perspective to some extent, in Chapter 4 when I argue for granting them negative rights.

My line of thought in this chapter is as follows: Firstly, we can consider the implementation of ethical boundaries from a robot perspective, but doing so means the consideration of granting moral status to robots and considering the actuality of robot consciousness[33], since moral patiency implies the ability to recognise self-worth[34], which implies some level of consciousness, at least in the sense of self-awareness. Rather, approaching the topic from an anthropocentric perspective, means focusing not on the actuality of robot consciousness, but only on the *perception* of robot consciousness and, as such, the *perception* of robot moral patiency[35]. This allows us to circumnavigate the seemingly intractable debate about consciousness, both in humans and robots.

Secondly, due to the possibility that we may perceive robots to be conscious and may interact with them as if they are socially intelligent (as explained in Chapter 1), I argue that the way in which we treat them may impact us socially and morally. I argue that, ultimately, treating them immorally may negatively impact the moral fibre of interactants. Thirdly, then – and going back to Turner (2019) – we should investigate ethical boundaries for human-sexbot interaction from an anthropocentric perspective because although sexbots have been designed and created for humans, interactants need protection from their own immorality (I explain what I mean by this below).

2. Robot moral status: perceived moral patiency

In reflecting on Turner's (2019: 170) statement that "protection in law often follows shortly after society has recognised a moral case for protecting something", I focus here briefly on

---

[33] Issues surrounding the actuality of robot consciousness were discussed in Chapter 1.

[34] "An entity has moral status if and only if it or its interests morally matter to some degree for the entity's own sake" (Jaworska & Tannenbaum 2018).

[35] The perception of robot consciousness was explained in Chapter 1.

the subject of the moral treatment of robots. Is there a moral case to consider in relation to our treatment of sexbots? Should they be treated well for their own moral sakes?

Taking on a robot perspective means that this 'something' that needs to be protected is the robot – they need protection from being treated immorally by interactants. In this capacity, we must then consider the possibility for robots to be moral patients (see e.g. Torrance 2013) in the sense that robots can be the target of immoral behaviour; that there is a possibility that they are "owed at least one duty or obligation" (Himma 2009: 21) by moral agents (being the human interactants) who direct their behaviour towards them.

Although this notion of robots being moral patients or, using Torrance's phrase, moral receivers, "is [at the] more speculative end of the [machine ethics] spectrum" (2013: 403) (the other side of the spectrum being the possibility for robots to be moral producers and also the side with which the majority of work concerning ethics of AI has concerned itself), these notions are "central to the wider question of the ethical significance of intelligent agents" (ibid.). This is because, as Floridi (2013: 135-136 as cited in Gunkel 2017) reminds us, "moral situations involve at least two interacting components – the initiator of the action or the agent and the receiver of this action or the patient".

Considering the moral patiency of robots, however, and considering whether they need to be protected from immoral behaviour, brings to light the question of whether robots can be morally wronged. Can they *feel what it is like* to be a moral patient and to be morally wronged? Can they suffer at the hands of those moral producers who treat them immorally?

In their chapter *The ethics of artificial intelligence* in *The Cambridge Handbook of Artificial Intelligence* (2014), Bostrom and Yudkwosky address the topic of machines having moral status. Drawing upon the definition of moral status from Kamm (2007), they state that:

"X has moral status = because X counts morally in its own right, it is permissible/impermissible to do things to it for its own sake" (Bostrom & Yudkowsky 2014: 321).

It is a commonly held belief that in order for something to have moral status, or be worthy of moral consideration, this something must be conscious in the phenomenological sense (Bostrom & Yudkowsky 2014; Jaworska & Tannenbaum 2018), because this would mean that they are able to subjectively experience suffering; that they can *feel what it is like* to be a moral patient that is treated immorally. This type of consciousness refers to "the capacity for phenomenal experience or qualia, such as the capacity to feel pain and suffer"

(Bostrom & Yudkwosky 2014: 321), and was discussed in more detail in Chapter 1. As such, in considering the moral status of robots, the issue whether robots are conscious, or will become, conscious – albeit in an artificial sense – arises, as well as the issue of whether this would warrant their being deserving of moral consideration (see e.g. Bryson 2009; Sparrow 2004; Torrance 2008; Turner 2019). This is because, as Singer (1977) states:

> "If a being is not capable of suffering, or of experiencing enjoyment or happiness, there is nothing to be taken into account. So the limit of sentience (using the term as a convenient if not strictly accurate shorthand for the capacity to suffer and/or experience enjoyment)[36] is the only defensible boundary of concern for the interests of others".

In the article *Artificial Agents and the Expanding Ethical Circle* (2013), Torrance considers expanding the 'ethical circle' (of moral patients and agents) so as to include artificial agents. His notion of such an expansion of the ethical circle pertains to questioning whether we should, just as we have included machine minds in our mental universe, also include machine minds in our moral universe (ibid.: 399). His article will be discussed in more detail in section 4 but, briefly, he questions whether the ethical circle should be expanded so as to include the moral treatment of artificial agents. What should our reasoning for this be? His consideration draws heavily upon the work of Singer (1977) as Singer "encouraged us to think beyond the parochial interests of the human race and to extend our ethical identification to other living species and to the wider environment" (Torrance 2013: 400). As stated above, for Singer (1977), phenomenal consciousness (or sentience), i.e. the capacity to suffer, is the key factor to take into account when considering the moral treatment of animals. In relation to the moral treatment of animals, Singer (1977) quotes Bentham (1982): "The question is not, Can they reason? nor Can they talk? but, Can they suffer?". If a being has the capacity to suffer at the hand of a moral agent who is treating them immorally, "there can be no moral justification for refusing to take that suffering into consideration" (ibid.).

However, taking the capacity to experience suffering as the condition to be treated morally well means assuming that the moral patient in question does actually have the

---

[36] Due to sentience being "the capacity for phenomenal experience or qualia, such as the capacity to feel pain and suffer" (Bostrom & Yudkowsky 2014: 322), the term can also be used to refer to phenomenal consciousness.

capacity to feel what it is like to suffer; that they are phenomenally conscious in that they experience the 'feeling what-it-is-like'. Such experience is, however, a private and subjective experience and, as such, is inaccessible to anyone else (see e.g. Nagel 1974). It cannot be accessed from the outside in its primary subjective givenness because nobody can experience it in its personal subjective form (Leube & Kircher 2003: 657).

Thus, if phenomenal consciousness is a subjective, first-person private state, how will we ever know if a sexbot has the capacity to suffer? As Singer (1977) states:

> "…how do we know if anyone, human or nonhuman, feels pain? We know that we ourselves can feel pain. We know this from the direct experience of pain that we have when, for instance, somebody presses a lighted cigarette against the back of our hand. But how do we know that anyone else feels pain? We cannot directly experience anyone else's pain, whether that 'anyone' is our best friend or a stray dog. Pain is a state of consciousness, a 'mental event', and as such it can never be observed. Behavior like writhing, screaming, or drawing one's hand away from the lighted cigarette is not pain itself; nor are the recordings a neurologist might make of activity within the brain observations of pain itself. Pain is something that we feel, and we can only infer that others are feeling it from various external indications".

As was discussed in Chapter 1, the topic of the actuality of robot consciousness is a thorny issue. As such, I put forward that we should rather circumnavigate the issue by way of focussing on *perceived* robot consciousness. Given the relation between phenomenal consciousness and moral patiency in the sense that being phenomenally consciousness means having the capacity to be a moral patient because one can *feel what it is like* to be morally wronged, we can consider this same relation in the context of *perceived* robot consciousness. Although sexbots may not *actually* be conscious and thus cannot *actually* be moral patients, there is the possibility that due to the *perception* of robots being conscious (due to our anthropomorphising them), human interactants may therefore *perceive* them as being moral patients the way in which humans are. Thus, putting forward that sexbots should not be treated immorally for their own sakes would be problematic because they are not *really* moral patients who can suffer at the hands of human interactants who may treat them badly. Rather, I put forward that due to the possibility of *perceived* moral patiency, human interactants

should not treat sexbots immorally for their own (human) sakes. The following sections will unpack this argument more.

3.   The negative implications of treating sexbots immorally

I here argue that due to the possibly that interactants may view their sexbots as being conscious (and thus perceive them as being moral patients) and bond with them in what appears to be a seemingly realistic way, treating them immorally may have a negative impact upon interactants themselves. I acknowledge that the concept of 'treating robots immorally' is referred to in broad terms throughout this dissertation. The very concept of what it would entail to treat a robot immorally is new territory as far as ethics and the law is concerned, and would require far more research and investigation than the confines of this dissertation would allow.

However, in the context of the arguments that I make here, my specific take on what the immoral treatment of sexbots entails, can be understood with regard to arguments made that sexbots are intended to be human simulacra (and this is elaborated upon in the sections that follow). As such, I put forward that at least in the context of the arguments I make here, my take on immoral treatment of a sexbot would be anything that we would consider as immoral treatment in relation to human-human interaction. I make this point by way of drawing upon arguments that sexbots are ultimately designed and created to be human simulacra and, therefore, interacting with a sexbot is symbolic of interacting with another human being (see e.g. Danaher 2017c). Therefore, in this sense, immoral treatment of a sexbot could be defined in terms of treatment that we generally consider to be immoral treatment of another human being.

I develop the argument that treating sexbots immorally would have negative implications for human interactants by using three sub-arguments: firstly, sexbots are more than mere inanimate objects to their human interactants due to the way in which we view and relate to them; secondly, the act of treating a sexbot immorally is morally abhorrent in itself due to it being symbolic of treating a human being immorally (even if the robot only mimics being human); and, thirdly, due to sexbots being more than mere objects (to their human interactants), and due to the consequent inherent immorality of treating them in immoral ways, I argue that treating a robot immorally may subsequently negatively impact upon the moral fibre of interactants.

3.1. Sexbots are more than mere objects

Although I am neutral for the purposes of this dissertation on whether or not sexbots are capable of possessing consciousness – particularly in the phenomenological sense as has been discussed in Chapter 1 – I argue that we cannot deem them as merely being inanimate objects. We cannot place social robots within the same group as any other object we utilise. This is because we do not view and relate to social robots the same way in which we view and relate to any other objects in the world. Although any type of bond with a robot may be unidirectional, and no type of reciprocation on the part of the robot truly indicates consciousness, the robot still does mimic reciprocation on a human social level (think back to Dautenhahn's (2007) definitions of social robots in Chapter 1), which impacts the humans with whom they interact. As such, I agree with Ramey (2005) that there may be a unique social relationship (albeit possibly unidirectional as far as genuine reciprocation is concerned) between a human and a social robot that is qualitatively different from the way in which we relate to any other object that we utilise (ibid.: 139).

We have more than a physical relation to them. Yes, one can have more than a physical relation to an inanimate object – children, for example, love their stuffed toys and it can be argued that these toys are created to elicit an emotional response from children. However, this type of interaction and emotional response differs from that which we experience with social robots since stuffed toys do not reciprocate emotion, whereas social robots do – even though this reciprocity may be mere mimicry. Given this, interactants may begin to see sexbots as being on the same plane as human beings (see e.g. Levy 2009). Therefore, although they may not actually be conscious, we may view them as being such, given the human-like way in which we are (or would be) able to relate to them (see e.g. Kanda et al. 2012; Melson, et al. 2009; Turkle 2006). Given this possibility, the superficial view to treat social robots (and thus sexbots) as mere objects does not seem viable – there is more to them than that – although *actually* granting them consciousness and considering them deserving of moral treatment the way humans are, may be taking it a step too far, especially given the contentiousness of the consciousness debate. This has all already been argued for in the previous sections and in previous chapters.

Given that I hold that social robots can be seen to be more than just any inanimate object due to the way in which we interact with them, I will now consider why the *act* of treating a sexbot immorally is wrong in itself. This is because not only may sexbots be viewed as being more than mere objects, but they can essentially be seen to be human

75

simulacra in that that they are being designed in our image, so as to facilitate the possibility for us to have human-like bonds with them.

3.2.   The act of treating a sexbot immorally is abhorrent in itself

Due to sexbots being created to foster the possibility for people to potentially *view* them as being conscious and on the same plane as human beings, sexbots may be said to ultimately be symbols of human beings. Therefore, any relationship formed with them is also symbolic of a relationship with a human being. Given this, one can argue that in treating a sexbot immorally, one is *symbolically* treating a human immorally, and this act can be seen to be morally abhorrent in itself.

This may seem like a leap, but it is important to then home in (again) on the humanistic aspect of these robots. They are specifically designed and created so that interactants will easily anthropomorphise them and relate to them in a humanistic social capacity. This is the whole point of their creation – to be human simulacra in every possible way, both physically and behaviouristically. As stated before, Kanda et al. (2004) have found a robot with a human-like body "causes people to behave unconsciously as if they were communicating with a human". Other studies have also confirmed the potential for interactants to attribute human-like aspects to robots and treat them as if they were human (see e.g. Kanda et al. 2012 & Levy 2007).

Sexbots are designed and created so that when an interactant physically – and emotionally – interacts with them, they are essentially performing an act which simulates the act that would be performed with another human being. This is for instance why moral questions arise regarding whether it would be wrong to allow a human to play out a rape fantasy using a sexbot as the victim. Both Sparrow (2017) and Turner (2019) ask this question. I hold the view that such an act would be immoral because the human-like form of the robot is intended to be symbolic of a human being, and moreover, if there is the possibility that an interactant may behave unconsciously as if they are interacting with a human, then playing out a rape fantasy with a robot simulates the enactment of an immoral act upon a human being and this act is immoral in itself (see also the discussion in Chapter 2).

Therefore, the act of treating a sexbot immorally is wrong in itself due to its symbolic meaning. If a human-like robot essentially symbolises a human being, and an interactant unconsciously behaves as if they are interacting with a human, yet treats this robot immorally, then the immoral act should be condemned. Due to the act itself being immoral,

there may be subsequent negative implications that may arise if interactants do treat their sexbots immorally. It is therefore important to address not only the morality of the act itself, but also consider how the act of treating a sexbot immorally may negatively impact interactants as moral beings.

3.3. Treating a sexbot immorally may negatively impact upon the moral fibre of human interactants

Given that we cannot deem sexbots to be mere objects due to the way in which we view and relate to them, and due to the act of treating a sexbot immorally essentially symbolising the act of treating a human immorally, there is the possibility that treating a sexbot immorally may negatively impact upon the moral fibre of interactants.

"[T]o treat androids as humans is not to make androids actually human, but it is to make oneself an expanded self" (Ramey 2005: 143) and the way we treat robots will affect ourselves and people around us. In light of this, David Levy, in his paper *The Ethical Treatment of Artificially Conscious Robots* (2009), argues that we should treat robots in the same moral way that we would treat any human, because not doing so may negatively affect those people around us "by setting our own behaviour towards those robots as an example of how one should treat other human beings" (ibid.: 214).

Similar questions have been raised as far as the moral treatment of animals is concerned. Turner (2019: 156) states:

> "If we treat animals with contempt, then we might start to do so with humans also. There is a link between the two because we perceive animals as having needs and sensations – even if they do not have the same sort of complex thought processes as we do. Essentially, animals exhibit features which resemble humans, and we are biologically programmed to feel empathy toward anything with those features".

If there is concern raised about the way in which we treat animals extending to the way in which we treat humans, then surely there should be even more concern regarding our moral treatment of sexbots who are realistic *human simulacra* as opposed to animals who may merely possess features that are exhibitive as human features?

As such, going back to Levy (2009), the main reason why he argues we should not treat robots immorally, is that if we take their embodiedness seriously, it would impact

77

negatively on our social relations with humans if we treated them immorally. This argument stems from the possibility that there is the potential for people to have meaningful interaction with them, leading to the formation of a bond on the part of the person who begins to *perceive* their sexbot as being sociable, intelligent and autonomous and, as such, being on the same plane as human beings, even if in fact, the interaction is unidirectional. This being the case, if we do begin to perceive sexbots as being on the same plane as human beings, Levy's (2009) argument that we should treat robots morally well, for our own sake, holds some weight.

One can, therefore, argue that since social robots (such as sexbots) are – in Levy's (2009) view – embodied computers, in treating a sexbot immorally, one is simulating the immoral treatment of a human being (as I have discussed above). If we do come to view these robots as being on the same plane as human beings, and yet not respect them as human beings, one can question theoretically whether this will lead to desensitising us towards immoral behaviour, thereby lowering the moral barriers of immoral acts. Would this potentially lead to human beings treating one another in such immoral ways?

Although such an argument can be likened perhaps to similar ones, for instance, debates about the impact of violent video games or pornography on society, the argument about sexbots differs in that "the nature of robots as three-dimensional entities capable of complex behaviours distinguishes them from other media" (Sparrow 2017: 467). Therefore, treating a sexbot immorally – by abusing them, or playing out a rape fantasy with them for instance – is more realistic than, say, video games, and, as such, is "more likely to encourage people to carry out the represented act in reality" (ibid.: 470). As Turner (2019: 159) states:

> "[S]imulating immoral or illegal acts with robots harms human society in some way, by condoning or promoting an unpleasant behaviour trait: an instrumental justification. This is a similar justification to the reason why cartoons depicting child pornography are often banned – even though no child was directly harmed in the process".

As such, I agree that our treating sexbots immorally may negatively impact upon the moral fibre of human interactants.

4. Considering ethical boundaries to human-sexbot interaction in order to protect human interactants from themselves

Given that the immoral treatment of sexbots may negatively impact upon the moral fibre of interactants, we may then consider the possibility of formulating and implementing ethical boundaries in relation to our interaction with them, specifically from an anthropocentric perspective. Thus, going back to Turner (2019), there is a moral case to protect something. This 'something' being the human interactant, and the moral case being that interactants need to be protected from their own immorality.

In his article *Artificial Agents and the Expanding Ethical Circle* (2013), Torrance discusses, as mentioned already section 2, from different ethical perspectives, the possibility of expanding moral consideration beyond humans or challenging the "anthropocentric ethical hegemony" (ibid.). Taking inspiration from Singer's *Animal Liberation* (1977), he considers the standpoints of anthropocentrism, infocentrism, biocentrism and ecocentrism on the topic of machine ethics and extending the moral circle beyond humanity, to be inclusive of AI.

Anthropocentrism and ecocentrism somewhat converge if we consider ecocentrism in relation to what is termed as the 'light green' position. Anthropocentrism is the view that defines ethics "as centred around human needs and interests: all other parts of the animate and inanimate world are seen as having little or no inherent value, other than in relation to human goals" (Torrance 2013: 404). 'Light green' ecocentrism has a similar outlook as anthropocentrism in that when it comes to environmental concern, concern must be driven by how any threats to the environment will also be a threat to human interests (ibid.: 409).

It is from a similar standpoint as these two ethical perspectives that I argue that we should show sexbots moral concern but that this moral concern should be for our own self-interests, and not necessarily for the interest of the robot. As Coeckelbergh (2010c: 210) states:

> "Robots with consciousness or the ability to 'demand their rights' seem to belong to the realm of science-fiction or at least the far future. Does this mean that current and near-future artificially intelligent robots should be excluded from our moral world entirely? Are there perhaps other ways of granting them moral consideration?".

As such, I agree that the ethical circle should be expanded so as to include sexbots – that they should not be excluded from our moral world entirely – but that we should consider including them from an anthropocentric perspective. Therefore, I take from Torrance (2013) that we should expand the ethical circle, yet contend that, in this case, we should not necessarily do so so as to challenge the "anthropocentric ethical hegemony".

Specifically, in relation to machine ethics, the anthropocentric view "sees machines, however intelligent or 'person-like' they may be, as being nothing other than instruments for human use" (Torrance 2013: 404). Similarly, the 'light green' ecocentrist view would only consider showing concern to the environment in as much as it is useful to us. Thus, if I formulate a 'light green' view in terms of ethics of social robotics, a similar standpoint is taken: only consider showing concern to a machine in as much as it is useful to us, or in as much as our interaction with the machine impacts on us. Neither view would take into consideration the moral status of the sexbot itself.

Generally speaking, the anthropocentric and light green ecocentrist views hold that humankind is the most important element of existence, as opposed to animals or the environment, for instance. Such a standpoint rightly appears to be problematic and particularly arrogant from animal ethics and deep environmental ethics perspectives, yet we cannot compare animals or the environment with the existence of sexbots. Unlike the animals with which we share the world, or the environment in which we live, sexbots have ultimately been created *by us, for us*. As Bryson (2009) argues, the fact of the matter is that "there would be no robots on this planet if it weren't for deliberate human decisions to create them" – the same cannot be said about animals, nor about the environment. Moreover, social robots (and sexbots in particular) have been created for us with the foresight that they have the potential to provide companionship for the disabled or elderly (Knapton 2017), and be used – albeit controversially – for good social ends, such as in the treatment of paraphilia (ibid.), or replace sex workers, thus potentially mitigating human trafficking (Levy 2007), or offer companionship to the aged or lonely (as has been discussed in the previous chapter).

As such, a sexbot can be understood to be a means to an end, and not an end in its own right that warrants moral consideration and, as such, there would be nothing wrong with taking on an anthropocentric perspective in relation to our moral treatment of them. Although such an instrumentalist view of sexbots may seem somewhat contradictory in relation to my argument in section 3.1. – that robots are more than mere objects – it must be emphasised that I hold that they are more than mere objects in the sense that we may *view* them as such. This

does not change what sexbots are in and of themselves – technological 'tools' "standing ready to serve the purposes of users" (Feenberg 1991).

Their being created by us and for us, however, does not mean, as argued above, that we should not show sexbots any moral concern – that we should treat them any which way we please – but rather that because they are useful to us in extremely intimate ways, and that the way in which we treat them may morally impact us, we should treat them well *for our own sakes*, so as to ensure that interaction with them remains beneficial to us or, in the least, not harmful. I thus take on an anthropocentric standpoint in relation to the moral treatment of sexbots, and when considering the subsequent ethical boundaries that we should consider in relation to their utilisation. This is at least until such a time as we have more definitive answers about artificial consciousness.

One could however contend that we still do not have definitive answers about human or animal consciousness, thus posing the question of why do we show humans and (some) animals moral consideration for their own sakes if we also do not know if they do actually experience suffering? Turner (2019: 155), however, argues that as far as the treatment of animals is concerned, and even humans for that matter, both human and animal rights have "intuitive appeal because it feels bad to watch another person [or animal] in pain". Moreover, "[u]nderstanding how others feel is one of humanity's most powerful tools for building value and belief systems which can bind society together" (ibid.: 155-156). Furthermore, Turner (2019: 156) also holds that "[f]ailing to protect the rights of other humans undermines the moral fibre of a community. It denies the basic emotional reaction we have to the perceived suffering of another. The same emotional reactions also govern our feelings towards animals, although to a lesser extent". As such, we need not concern ourselves with the actuality of consciousness because if we perceive the moral patient as being negatively affected by immoral treatment – such as by acting as if they are suffering – we must not deny our empathetic response because doing so may negatively impact our moral fibre.

Moreover, going back to Turner (2019) again, there is a moral case for protecting humans from themselves in that moral standards of human interactants, and society in general, need to be protected. Therefore, we should take on an anthropocentric perspective when considering the implementation of ethical boundaries for human-sexbot interaction because we must ultimately implement them so as to protect humans from themselves – from their own immoral behaviour that may impact their moral standards and the moral standards of society as a whole. Boundaries must be formulated so that this technology remains beneficial to human interactants.

81

The argument for considering ethical boundaries in relation to human-sexbot interaction from an anthropocentric perspective is a strong one, regardless of whether sexbots are, or are not conscious, or if they may one day become conscious – whether we will be able to prove this consciousness or not. This is because if sexbots are not conscious, then we must do as I argue we should do and formulate and implement ethical boundaries for human interaction with them for the sake of interactants due to their viewing and treating sexbots as if they are conscious. And if sexbots are – or come to be – conscious? Even if in a new 'artificial' sense? We may then consider the possibility that they could demand to be treated morally well. As Asaro (2006: 12) speculates:

> "… robots might simply demand their rights. Perhaps because morally intelligent robots might achieve some form of moral self-recognition, question why they should be treated differently from other moral agents … This would follow the path of many subjugated groups of humans who fought to establish respect for their rights against powerful sociopolitical groups who have suppressed, argued and fought against granting them equal rights."

However, were we to treat sexbots morally well for our own sakes, then this would simply mean that sexbots would reap the benefits of our treating them well and implementing ethical boundaries from an anthropocentric perspective anyway, for they would then be treated morally well (as they would then deserve to be) and interactants would still be treating them well for their own sakes. Such a scenario of robots demanding moral consideration may then not arise. Either way, implementing ethical boundaries from an anthropocentric perspective would be beneficial. If we take on the anthropocentric perspective, then it essentially does not matter if sexbots are or are not conscious. The argument becomes redundant because all that matters is how interactants view, treat, and relate to these robots. Thus, the importance of considering ethical conditions (in the form of ethical boundaries) for human-sexbot interaction – even from an anthropocentric perspective only – would still hold.

5. Conclusion

The nature of sexbots (their human-like appearance, and ability to socially interact with us, albeit in limited ways) means that human interactants can interact with them in extremely intimate ways, similarly to the way in which one may interact with a human companion. As

such, it is important for us to consider the way in which human interactants treat their sexbots, and how the way in which they treat them may morally impact both the human interactant, and society at large.

In this chapter, I ultimately argued that due to the possibility for human interactants to perceive sexbots as being phenomenally conscious, there is the possibility that human interactants may then perceive sexbots as being moral patients in the sense that, due to their human-likeness, it may *seem as if* they can feel what it is like to be treated immorally. I then argued that due to this perception, the moral fibre of human interactants would be damaged should they treat sexbots immorally. This was argued in three sub-arguments: firstly, sexbots are more than mere objects due to the way we perceive them; secondly, the act of treating a sexbot immorally is immoral due to the way in which we perceive them; thirdly, due to points one and two, treating a sexbot immorally would morally harm human interactants. Given that our immoral treatment of sexbots may morally harm *us* (it cannot harm sexbots because they are not *actual* moral patients), I then put forward that we should consider the formulation and implementation of ethical boundaries in relation to human-sexbot interaction that would essentially prevent us from treating sexbots immorally so as to protect *us* from our own immoral actions towards sexbots.

Drawing upon the formulation and implementation of ethical boundaries in relation to our interaction with sexbots, in the following chapter, I will argue that we could consider granting sexbots negative rights. Considering the granting of negative rights to sexbots can be seen as a practical way to implement my suggestion that human interactants should not treat sexbots immorally, and that formulating and implementing ethical boundaries in relation to human interaction with sexbots is necessary.

Chapter 4

Granting sexbots negative rights to protect the moral fibre of human interactants

1.   Introduction

In Chapter 3, I put forward the argument that, generally speaking, human interactants should not treat sexbots immorally because doing so could negatively impact upon the moral fibre of the human interactant, and thus, potentially, the moral fibre of our society, if many people make use of sexbot technology. As such, the argument against the immoral treatment of sexbots was analysed from an *anthropocentric perspective*: human interactants should not treat sexbots immorally – not for the sake of sexbots, but for their (the human interactants) own moral sakes.

Of course, as far as technological development is concerned, we generally hope that advances will be beneficial to society at large. Such is the case with AI and there is widespread agreement that it is vitally important to ensure that AI is beneficial to us[37]. However, given how nuanced the ethics of human interaction with sexbots is, how can we try to ensure that this technology is beneficial to human interactants, and to our society at large?

One way is to understand that it is the human interactants who are essentially responsible as to whether their interaction with sexbots is beneficial or not, hence my argument that we consider the implementation and formulation of ethical boundaries for human-sexbot interaction from an anthropocentric perspective. As has been discussed thus far, we cannot seek to protect society from moral harm by taking on a 'robot' perspective because for now the issue of robot consciousness is too uncertain. Moreover, if we regulate the interaction with sexbots too far or stop their development, although we may protect human interactants and society from potential harm, it would also mean that human interactants and society would either not benefit from the technology at all, or at least would not experience their full range of potential benefits. This would mean not only restricting the

---

[37] Refer to Footnote 3.

potential negative moral impacts upon human interactants and society, but potential beneficial impacts too (as was also discussed in Chapter 3).

Do we, then, find ourselves in a moral quandary, or ethical dead-end, as far as the beneficial use of sexbots is concerned? Not necessarily, if we approach the issue from an anthropocentric perspective once again. As the Future of Life Institute (FLI) (https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/?cn-reloaded=1) states:

> "[O]ur civilization will flourish as long as we win the race between the growing power of technology and the wisdom with which we manage it. In the case of AI technology, FLI's position is that the best way to win that race is not to impede the former, but to accelerate the latter, by supporting AI safety research".

As such, we should not impede upon the development of sexbots, nor impede upon the roles that they can fill in society. Rather, in this chapter, the argument is that we should focus on humans and the way in which we interact with sexbots (i.e. *how* we use them) so as to manage our interaction with sexbots in such a way that society is maximally benefitted. This draws upon arguments made in Chapter 3 that sexbots should not be treated immorally for our (human interactants) moral sakes. Approaching the issue of how humans treat robots from an anthropocentric perspective means finding a way to try to ensure that human interactants do not treat sexbots immorally regardless of what they are used for – as carebots, as replacements for sex workers, or as life companions – since treating them immorally may negatively impact upon the moral fibre of human interactants and society at large. This chapter suggests that considering granting negative rights to sexbots could be one way in which we can go about the difficult task of establishing such conditions for human interaction with sexbots.

This may not be a wholly perfect approach in the sense that, for example, Chapter 2 put forward how treating a sexbot immorally (such as by using it to enact a rape scene) may be beneficial as a form of sex therapy for sex offenders. However, that kind of context is unique and controlled, and given the pressure on ensuring ethical considerations stay ahead of technology, I think that it is a generally informative approach to take for now. Moreover, generally speaking, should we not act morally well in all our endeavours as civilized beings? Regardless of what is at the receiving end of our moral actions? After all, there is a

85

longstanding "anthropocentric tradition in ethics that is exclusively agent-oriented, no matter the wide range of incompatibilities that exist, for example, between virtue ethics, consequentialism, and deontologism" (Gunkel 2012: 99). Treating sexbots immorally as a method of catharsis which would benefit certain human interactants (namely sex offenders) is a very particular instance and if, generally speaking, there is consensus that we should act morally in all our endeavors as human moral agents, then I hold that such a particular instance wherein immoral treatment may be beneficial should not weaken my argument that, on the whole, abstaining from treating sexbots immorally would be more beneficial to human interactants and our society than not doing so.

Given this, how do we then *regulate* the way in which human interactants interact with sexbots? I put forward that a way in which we can try to ensure that human interactants treat sexbots morally well is to establish ethical boundaries for our interaction with them in the form of granting sexbots *negative rights*. Coeckelbergh (2010c: 210) notes that the issue of whether we should take artificially intelligent robots into moral consideration at all is often framed in terms of rights, and the topic of robot rights is an ethically loaded one that has, over the past few years, gained considerable attention.

My consideration of granting negative rights to sexbots is important, not only because I hold that it is a possible, although tentative, suggestion as a solution to an issue that this dissertation raises – namely, how do we ensure that human interactants do not treat sexbots immorally? – but also because it makes a worthwhile contribution to current literature on robot rights for three reasons: firstly, it grapples with both the descriptive and normative aspects of the topic, which seems to often be amiss in current machine ethics literature – an issue that Gunkel (2017) has raised. In other words, dealing with the questions of both *can* robots have rights, as well as *should* robots have rights. Secondly, it deals with the notion of robots as moral patients, which has seldom been discussed in current literature. Thirdly, in current literature on robot rights, few authors clarify which *kind* of rights should, or should not, be granted to robots – something with which Tavani (2018) takes issue. Although homing in on the concept of negative rights remains a broad account of the kind of rights that should be granted, it is, nonetheless, a distinction I have not yet come across, and a first step towards a finer account of robot rights. While my focus is specifically on sexbots, however, given that sexbots are social robots, my arguments can be applied in the context of considering the granting of negative rights to social robots in general.

This chapter will, firstly, clarify the moral positions that human interactants and sexbots have in comparison to, and in relation to, one another in terms of human moral

patiency. It will, firstly, discuss how robots may be conduits of human moral action towards human moral patients and, secondly, how humans can be the moral patients of the moral actions of robots. I then put forward a third moral relation we may consider: human interactants as moral patients of their own agential moral actions towards robots. Given that human interactants can be the moral patients of their own agential moral actions towards robots (and thus can be morally harmed should they treat sexbots immorally), I then go on to consider granting negative right to sexbots so as to prevent human interactants from treating sexbots immorally, therefore protecting human interactants' moral fibre. I, firstly, discuss the very notion of granting rights to robots and whether this is a possibility before, secondly, considering why granting negative rights to sexbots may be a way to ensure that sexbots are beneficial to society.

2. Human interactants as both agents and patients of their moral actions

Before embarking upon my argument for granting negative rights to sexbots, we need to firstly clarify the moral positions that human interactants and sexbots have in comparison to, and in relation to, one another. "The concept of moral agency should be distinguished from that of moral patiency. Whereas a moral agent is something that has duties or obligations, a moral patient is something owed at least one duty or obligation" (Himma 2009: 21). In particular, "the behavior of moral agents is governed by moral standards" (ibid.) and, therefore, moral agents are held morally accountable for their actions. Further, "[a]dult human beings are, for example, typically thought to be moral agents" and "[m]oral agents are usually, if not always, moral patients; all adult human beings are moral patients" (ibid.). Humans are thus both *actual* moral agents and patients.

Sexbots, however, are *perceived* as being moral patients by the humans that interact with them (this argument was made in Chapter 3). As far as sexbots as moral agents is concerned (and robots in general), this is a hugely contentious subject, full discussion of which would go beyond the confines of this dissertation and linking to the robot consciousness debate (introduced in Chapter 1). Moreover, the consideration of the possibility for sexbots to be moral agents does not play as big a role in terms of my argument to follow and, as such, does not require detailed discussion in the context of this chapter.

My focus turns away from the topic of moral agency, and towards moral patiency given the general lack of attention it has received in robot ethics ('roboethics') literature (see

87

e.g. Gunkel 2012 & Levy 2009). Referring to Levy's (2009) literature review, Gunkel (2012: 97) notes that "… roboethics has been exclusively focused on questions regarding both human and machine moral agency". In Chapter 3, the notion of sexbots being *perceived* as moral patients because of the tendency for human interactants to *perceive* them as being phenomenally conscious – due to anthropomorphising them – was discussed. As such, I will not need to spend more time on the topic in this chapter. I will, therefore, shift focus away from a robot perspective as far as being a moral patient is concerned and, once again, towards an anthropocentric one, i.e. focusing on human interactants as being moral patients with regard to their interaction with sexbots. Specifically, in this section, I will put forward my reasoning for why *human interactants are the moral patients of their own agential moral actions towards sexbots.*

An anthropocentric perspective of moral patiency in relation to human interaction with robots is a nuanced perspective, and there are two particular perspectives I have come across within this category on which I would like to focus: Firstly, humans being moral patients to moral actions conducted by other human agents through the use of a robot (or computer) (Gunkel 2012: 96) and, secondly, the issue of humans being moral patients to the moral actions of robots themselves – which then brings up the issue of machine moral agency. This distinction is succinctly made by Gunkel (2012: 101):

> "The first is computer ethics, which is concerned … with questions of human action through the instrumentality of computers and related information systems. In clear distinction from these efforts, machine ethics seeks to enlarge the scope of moral agents by considering the ethical status and actions of machines".

Following a discussion of these two categories, I will then put forward a third distinction of human moral patiency in relation to human interaction with robots: human interactants as moral patients of their *own* agential moral actions towards robots.

This section considers human moral patiency in relation to robots in broad terms (section 2.3. is more focused on social robots due to discussion on perceived moral patiency), and not specifically in the context of sexbots. However, these generalized arguments are applicable in the context of human-sexbot interaction too as sexbots are social robots, as pointed out often now.

2.1. Robots as conduits of human moral action towards human moral patients

Although this category of human moral patiency is related to computer ethics, it can also be applied to robot ethics. Regarding this first distinction, computer ethics, for example, "endeavors to stipulate the appropriate use and/or misuse of technology by human agents for the sake of respecting and protecting the rights of other human patients" (Gunkel 2012: 96). There is actually a "Ten Commandments of Computer Ethics" published by the Computer Ethics Institute in 1992 wherein each commandment specifies what would constitute as appropriate or inappropriate use of computers in this regard. The first commandment, for example, states: "Thou shalt not use a computer to harm another person". As such, we can say that computers are ultimately deployed by humans, used for a human purpose and, as such, have an effect on humans. An example could be using computer technology through social media to spread fake news or deface somebody's character.

Or, as far as robotics is concerned, there is the possibility of directly commanding a robot to injure another human being. In such instances, a human agent would not be directly interacting with another human patient, they would be treating a human patient immorally through the use of technology – such as a computer or robot; technology would be the conduit of immoral action on behalf of the human agent, directed at another human patient. Although the robot is conducting the immoral action upon the human moral patient, the difference (as compared to the second perspective discussed below) is that there is direct human intervention whereby the moral decision is ultimately made by a human, but the human agent uses technology to then inflict the moral harm that is the result of the decision they have made. For example, in terms of autonomous weapons systems (AWSs), there is a distinction between AWSs which "operate entirely independently of human controllers, and tele-operated unmanned weapons systems, which are still under remote human control" (Danaher, et al. 2017: 61-62). Tele-operated weapons systems would be a case of a human agent ultimately making a moral decision as to whether to harm a human moral patient or not, but using an AWS to carry out the decision. AWSs that operate entirely independently of human controllers would fall under the second category relating to machine ethics – humans being direct moral patients of moral decisions made by technology or, particularly in this instance, robots. This brings us to the next section.

2.2. Humans as moral patients to the moral actions of robots

As far as the second perspective is concerned, machine ethics (ME), for example, "seeks to enlarge the scope of moral agents by considering the ethical status and actions of machines" (Gunkel 2012: 101). It (ME) "reasserts the privilege of the human and considers the machine only insofar as we seek to protect the integrity and the interests of the human being" (ibid.: 102). It considers the possibility of machines to be guided by ethical principles in the decisions that it makes about possible courses of action (Anderson & Anderson 2007: 15). As such, the machines in question are machines that make decisions and act autonomously (without human intervention) by way of "[combining] environmental feedback with the system's own analysis regarding its current situation" (ICRC 2019: 7). Given this understanding of autonomous decision making systems (ADM systems) that have the potential to be moral agents, we can then consider the possibility that humans can be moral patients to the moral decisions and actions of AI. Specifically, in our context, this potentiality means that robots could harm humans.

The topic of the possibility for machines to be considered moral agents is a broadly contested and complicated one, full discussion of which would go beyond the confines of this dissertation. However, it is worth noting some arguments that have been made concerning the topic, and Tavani (2018) gives a concise overview of some of the arguments made. Generally speaking, the topic is one which questions whether machines can be moral agents – is morality programmable? – and what conditions they would have to fulfill in order to be considered moral agents, as well as the impact that these agents would have on us. As Behdadi and Munthe (2020: 3) question: "Which conditions, if any, are necessary and sufficient for an artificial entity to be a moral agent (i.e. an AMA)?". Tentative answers abound: Floridi (2013), for example, holds that some autonomous artificial agents can be moral agents because they are "sources of moral action". Johnson (2006), however, holds that artificial agents are not moral *agents* but are rather moral *entities* because although artificial agents can be autonomous, they do not possess true freedom. Then there is Himma (2009) who contends that artificial agents cannot be moral agents given that they are not conscious and lack intentionality[38].

---

[38] In philosophy, intentionality can be understood as "the power of minds and mental states to be about, to represent, or to stand for, things, properties and states of affairs. To say of an individual's mental states that they have intentionality is to say that they are mental representations or that they have contents" (Jacob 2019).

Other well-known researchers weighing in on the issue include Asaro (2006), Bostrom and Yudkowsky (2014), Brundage (2014), Deng (2015), Lumbreras (2017), McDermott (2008), Moor (2006), Sullins (2006), Torrance (2013), van Wynsberghe & Robbins (2019), Wallach and Allen (2009), Wang & Siau (2018), and many others.

Different sets of conditions for moral agency are suggested: A combination of free will, consciousness, and moral responsibility (Wallach & Allen 2009); a combination of the abilities to be interactive, autonomous, and adaptable (Floridi & Sanders 2004), and a combination of autonomy, responsibility and intentionality (Sullins 2006). Furthermore, consider whether we need to ensure AMA's are both ethically productive and ethically receptive (Torrance 2013), or is the ability for rational deliberation all that is needed (Abney in Lin et al 2012: 47)?

Although it is debatable whether robots can or cannot truly be moral agents given how philosophically loaded the topic is and its obvious relation to consciousness debates again, it remains that, regardless of this uncertainty, humans can still be moral patients of the actions of autonomous machines that act without direct human intervention. For instance, and going back to the example mentioned above of AWSs, although we could debate endlessly about whether an AWS that acts without human intervention is a moral agent, the fact remains that regardless of its status as a moral agent, it seems it still ultimately makes the moral 'decision' to kill a civilian or not, and this civilian would be the moral patient of this moral 'decision' – whether they lived or died, or whether the AWS had full or limited moral agency.

Thus, as stated above, there is the potential for human beings to be harmed by ADM technology. Such a possibility could also arise in the case of a self-driving car having to decide whether it should hit and kill a pedestrian crossing the road or veer off the road to avoid killing the pedestrian, but potentially killing the passenger of the car. This is not to say that were the car to kill the pedestrian or the passenger, it would hold full moral responsibility for their death – this is another complex issue entirely[39] – nor is it to say that the car is, in and

---

[39] The topic of moral responsibility is also a contentious one and there remains what can be termed a *responsibility gap* when it comes to who should be held responsible for the actions of autonomous systems – such as self-driving cars or AWSs. The responsibility gap can be explained as such: "Traditionally we hold either the operator/manufacturer of the machine responsible for the consequences of its operation, or 'nobody' (in cases, where no personal fault can be identified). Now it can be shown that there is an increasing class of machine actions, where the traditional ways of responsibility ascription are not compatible with our sense of justice and the moral framework of society because nobody has enough control over the machine's actions to be able to assume the responsibility for them. These cases constitute what we will call the responsibility gap"

91

of itself, a moral agent. Rather, it is to say that moral responsibility and agency aside, the human who was killed would have been killed due to a decision ultimately made by the car (although the groundwork for the decision would be based on its programming). At that moment, there is no direct human intervention wherein a human is making the decision to hit the pedestrian, or potentially kill the passenger.

2.3. Human interactants as moral patients of their own agential moral actions towards robots

There is, however, also a third anthropocentric perspective that we can consider. This is the perspective of a human interactant being a moral patient of their own agential moral actions towards a robot (and, in this context, particularly a social robot and, thus, a sexbot). That is, the impact of the very action taken by a human interactant (who is a moral agent) towards a robot – which is a *perceived* moral patient – is redirected towards the human agent, making them a patient of their own immoral actions because their moral fibre is impacted by the way in which the robot is treated. Moral patiency – as mentioned in Chapter 3 and earlier in this section – can be understood as the case of being a target of moral action. In this instance, human interactants would not be *direct* targets of their own actions, but rather *indirect* targets – like a bullet ricocheting off its direct target and injuring an innocent bystander who becomes an indirect target of the shooter. They (human interactants) are indirectly impacted by way of their moral fibre being negatively impacted in the case of their direct immoral treatment of robots.

In this case, the robot is the direct target of the immoral treatment – and the *perceived* moral patient – while the human interactant is the indirect target – and the *actual* moral patient. As such, humans are indirect recipients of immoral action because robots cannot *actually* be recipients. Robots are not *really* impacted (for now leaving aside the possibility of robot phenomenal experience and consciousness (Chapter 1), which, if it comes to pass, would of course add a layer of the robot as moral patient to this discussion) – we (the human

---

(Matthias 2004: 177). As stated by Elish (2019: 41): "Questions having to do with the responsibility of various agents in complex, computational systems are not new. Such issues have been looked at from the diverse perspectives of law and policy (Calo, Froomkin and Kerr 2016), human factors of engineering (Cummings 2006; Sheridan and Parasuraman 2005), systems design (Friedman 1997; Leveson 2011), ethics (Lin, Abney, and Bekey 2014; Coeckelbergh 2011; Bryson, Diamantis and Grant 2017) and the sociology of risk and innovation (Perrow 1984; Vaughn 1996)".

interactants) are. Moreover, Danaher (2019: 132) notes a moral patient as "a being who possesses some moral status – i.e. is owed moral duties and obligations, and is capable of suffering moral harms and experiencing moral benefits – but who does not take ownership over the moral content of its own existence". As far as human interactants being moral patients of their own moral actions is concerned, referring to Danaher's (2019) definition, human interactants can suffer and experience moral harms and benefits of their own agential actions: specifically, moral harms by way of their moral fibre being negatively impacted is an example of this kind of suffering.

Interestingly, Danaher (2019) actually argues that the rise of robots could bring about a decrease in our own moral agency: "That is to say, [the rise of robots] could compromise both the ability and willingness of humans to act in the world as responsible moral agents, and consequently could reduce them to moral patients" (ibid.: 120). For example, and as elaborated upon in Danaher's (2019) article, consider an instance in which someone spends all their time with their sexbot. As a consequence, the human interactant loses motivation to do anything of real consequence – go out and meet new people, or spend time with a human partner – because it takes more effort. As such, this human interactant can spend all day at home, enjoying all the pleasure they desire (ibid.: 130-131). As Danaher (2019: 131) states:

> "[T]he rise of the robots could lead to a decline in humans' willingness to express their moral agency (to make significant moral changes to the world around them). Because they have ready access to pleasure-providing robots, humans might become increasingly passive recipients of the benefits that technology bestows".

This is a compelling argument and worth consideration (more on issues related to this in Chapter 5). However, I rather argue here not so much that our moral agency could itself 'decrease' due to our interaction with sexbots, but rather that our moral agency could be negatively impacted in the sense that as moral agents, our moral fibre may be negatively impacted should we mistreat sexbots, thus causing us, as moral agents, to possibly act immorally towards other human beings with whom we share the world, and towards ourselves.

Most ethics are agent-oriented – hence Floridi & Sanders (2004) refer to this orientation as the "standard" approach. As such, a patient-oriented approach is "non-standard" – "it focuses attention not on the perpetrator of an act but on the victim or receiver

of the action" (Gunkel 2012: 108). Considering the possibility of human interactants being both agents and patients in a given instance bridges such a divide between a standard and non-standard approach. This is because human interactants – as moral agents – have the capacity to treat robots in moral or immoral ways. However, such treatment indirectly impacts human interactants as moral patients – they are, too, indirect receivers or victims of their own moral actions given that treating a robot immorally may negatively impact upon their own moral fibre.

Although in this section I have generally made my arguments with reference to robots in general and not sexbots in particular (this is due to trends in the literature that I have referenced), the same arguments would be appropriate in the context of our interaction with sexbots by virtue of them being social robots.

3. Understanding the concept of 'rights' and the possibility of granting them to robots

The term 'rights' is used in various contexts. Rights can be understood as "entitlements (not) to perform certain actions, or (not) to be in certain states; or entitlements that others (not) perform certain actions or (not) be in certain states" (Wenar 2020). Examples of the various contexts in which the term is used include animal rights, and of course, human rights, such as the right to life, the right to healthcare, the right to privacy, to name only a few. As Wenar (2020) states: "We encounter assertions of rights as we encounter sounds: persistently and in great variety". To make sense of the various ways in which different rights can be asserted, they can be grouped together into four categories based on common attributes:

> "*Who* is alleged to have the right: Children's rights, animal rights, workers' rights, states' rights, the rights of peoples.
>
> *What* actions or states or objects the asserted right pertains to: Rights of free expression, to pass judgment; rights of privacy, to remain silent; property rights, bodily rights.
>
> *Why* the rightholder (allegedly) has the right: Moral rights are grounded in moral reasons, legal rights derive from the laws of the society, customary rights exist by local convention.

How the asserted right can be affected by the rightholder's actions: The inalienable right to life, the forfeitable right to liberty, and the waivable right that a promise be kept" (Wenar 2020).

Turner (2019) brings to the fore two aspects of the nature of rights that are important in relation to the idea of granting rights to robots: rights are *social constructs* and rights are *fictions*. Regarding the former – rights being social constructs – Turner (2019) explains this by way of drawing upon the work done by Wesley Hohfeld who put forward four basic components of rights termed "the Hohfeldian incidents" which are: privileges, powers, claims and immunities (Turner 2019 & Wenar 2020). Moreover, as Turner (2019: 134) explains, in addition to unpacking these four components of rights, "Hohfeld's other key insight was to pair each category of right in a reciprocal relationship with a right held by another person. The four categories listed above correspond to the following: duty, no-claim, liability, disability. Thus, if Person A has a claim to something, Person B must have a liability to provide Person A with that thing". Wenar (2020) explains this more formally. Looking at each category of rights, each reciprocal relationship can be explained as follows:

Privileges: "A has a privilege to φ if and only if A has no duty not to φ".

Claims: "A has a claim that B φ if and only if B has a duty to A to φ".

Powers: "A has a power if and only if A has the ability to alter her own or another's Hohfeldian incidents".

Immunities: "B has an immunity if and only if A lacks the ability to alter B's Hohfeldian incidents".

According to Turner (2019: 135), this reciprocal dynamic highlights how rights are social constructs: "[t]he correlatives to each right show that they do not exist in a vacuum. Rather, they are held against other people or entities". As such, it would not make much sense for me to claim my right to freedom of speech if I lived on a deserted island where there was no threat to my freedom of speech being taken away from me. Given this understanding of rights as social constructs, it then becomes apparent as to why we should consider rights in relation to AI and, more specifically, potentially granting rights to social

robots and, in this dissertation in particular, sexbots: we will be living alongside them and interacting with them in 'reciprocally'[40] human-like ways.

Turner (2019: 135) adds that "[t]o hold rights, therefore, is to coexist with others capable of upholding or infringing upon those rights". Given the current rapid development in the field of AI and robotics, we can begin to analyse such a coexistence with AI and robots in particular. Take the right to life, for example, in the context of AI warfare[41]: an autonomous robot can uphold, or infringe upon, a human right to life in a situation where, without direct human intervention, it had to decide whether to aim and kill a civilian or not. Although we could not necessarily hold the robot responsible for the killing of the civilian[42] – as discussed above, the topic of moral responsibility in such situations is a complex one – the existence of such technology that can directly impact human lives so dramatically would still be something with which humans would have to coexist.

We could also particularly consider the way in which a sexbot may infringe upon our rights. Given the social aspects of these robots, there could be a variety of ways in which our interaction with them and the technology itself may morally impact us given the potential ability to speak and act in human-like ways, and its machine learning capabilities. We may have a greater understanding of the ways in which sexbots may directly morally impact us as the technology develops, but let us consider one futuristic example. We, as human beings, have the right to privacy. A right to privacy can be understood in many different contexts,

---

[40] I place 'reciprocally' in inverted commas because it remains the case that, presently, we do not have *genuinely* reciprocal relations with sexbots given that they cannot genuinely reciprocate human emotions. Granted, this may not always be the case, but it seems it will at least remain the case for the near future.

[41] Although my focus is on sexbots, when it comes to the way in which *they* (sexbots) may morally impact *us*, we could only speculate at this point given that this is technology that, although is developing rapidly, has not reached a point in its development where its own decisions could have an actual impact on us. Thus, there is no empirical work on which I can draw as noted in previous chapters too. For now, I draw often upon examples of AWSs when it comes to elaborating upon the way in which robots could directly morally impact us, because, given that its intention would be to kill human beings, it is a clear example to use to exemplify my arguments of how humans could be morally impacted by decisions made by AI systems and, specifically artificially intelligent robots. However, it must be noted that lethal autonomous weapons also do not exist (at least legally) as yet, although literature on the topic abounds.

[42] The consideration of who should be responsible in a situation where an autonomous weapon kills someone is a contested subject. As Davey (2016) states on the Future of Life Institute website: "Intention must be demonstrated to prosecute someone for a war crime, and while autonomous weapons may demonstrate outward signs of decision-making and intention, they still run on a code that's just as impersonal as the code that glitches and freezes a computer screen". This relates to the issue of moral responsibility as mentioned in section 2.2.

96

however, specifically in the digital age in which we live, "detection devices track our movements, our preferences, and any information they are capable of mining from our digital existence. This data is used to manipulate us, rob from us, and engage in prejudice against us – at times legally" (Goldstein, et al. 2018). Scheutz (2012) raises concern about the possibility for companies to manipulate human interactants through the use of social robots: "For example, a company might exploit the robot's unique relationship with its owner to make the robot convince the owner to purchase products the company wishes to promote". In the context of sexbots with whom we may interact in extremely intimate ways, a company manipulating your intimate bond with your sexbot would be controversial and, moreover, an invasion of privacy. Thus, similarly to the AWS example, although the sexbot would be the conduit of this action of the company towards a human interactant, it remains the case that this is technology with which we would have to coexist and that would add to increasing concern in terms of our right to privacy being infringed upon in an increasingly digital age.

Moreover, and returning to the concept of the right to life, we could even consider the possibility that a human may infringe upon a robot's 'right to life' if we did ever reach a point where we could consider robots to be consciously aware. Even though robots are not currently conscious, our coexistence with them and the tendency to anthropomorphise social robots in particular possibly associates our turning them off, or disposing of them, as 'killing' them, and studies have analysed the reactions of human interactants thereof (see e.g. Horstmann, et al.: 2018). Although we have not reached such a point yet, as always, the nature of ethics of AI calls upon its researchers to consider such future possibilities and the moral ramifications thereof.

A second aspect of the nature of rights (connected to rights as social constructs) that Turner (2019) discusses is rights as fictions. Turner (2019: 135) states: "[rights] are communal inventions which do not have any independent, objective existence beyond our collective imagination". As such, depending on various contexts in which they may be considered, their form can be shaped differently – they are malleable fictions. After all, we need only look at the history of rights to understand their inherent malleability: progress made in human rights as far as slavery and women's rights are concerned, for example, as well as the advent of the protection of animal rights from the 17th century onwards (see e.g. Turner 2019: 140). Given the progress of rights due to their malleability, it seems a natural progression to consider robot rights as the technology advances and becomes more embedded into our social lives.

97

Turner (2019: 136) notes that "some rights are treated as more valuable than others, and belief in them may be more widely shared, but there is no set quota of rights which prevents new ones from being created and old ones from falling into abeyance". Understanding the fictitious nature of rights is important in relation to granting rights to robots given the instinctively negative reaction that some have to the topic (ibid.: 137). Such negative reactions sometimes stem from approaching the topic from the view of the "property account" (see e.g. Coeckelbergh 2010c) which holds that "a robot would be eligible to be granted rights, or at least some level of moral consideration, if it could convincingly show that it possessed one or more essential properties, such as consciousness, intentionality, rationality, personhood, autonomy, sentience, etc. (Tavani 2018: 6)[43].

This view, however, is problematic given that we can barely concretely prove that human beings possess such properties, yet there is no doubt that humans deserve human rights. As such, how can we use the property account as a determining factor for whether robots should be granted rights or not? As Coeckelbergh (2010c: 212) notes:

> "[It] is difficult to agree which ontological property is morally relevant and to provide proof that [any] entity in question has that property. Our moral intuitions differ on what criteria are the relevant ones. Moreover, even if we could agree on that, then can we provide 'hard' evidence for the presence of these criteria in a particular robot, given that the criteria are rather abstract?".

Such a view, therefore, takes for granted that humans possess such properties, making them eligible to be possessive of rights. Robots however, in not possessing these properties, on this view cannot be considered to deserve rights. Turner (2019: 137), however, holds that such a view may be a result of people having an "unspoken assumption that rights are a fixed quantity, like unchanging commandments written on tablets of stone". As such, sceptical reactions about granting rights to robots may be due to such an unspoken assumption.

Moreover, specifically as far as legal personhood in relation to the law is concerned, considering robots to be legal persons upon which we could confer rights is not controversial from a legal perspective. "To be a legal person is to be the subject of rights and duties. To

---

[43] Of course, such a conception of rights is concerned with rights that we, by virtue of being human, possess i.e. human rights. Other rights, such as animal rights or property rights, exist that are not founded on such properties. This is a philosophical approach to the possibility of robot rights, which differs from the legal approach to the matter, which will be discussed below.

confer legal rights or to impose legal duties, therefore, is to confer legal personality...'' (Smith 1928: 283). From a metaphysical or ethical perspective, the notion of considering robots to be legal persons is a philosophically loaded and controversial one. However, as stated by Bryson et al. (2017: 277):

> "Legal personality is not necessarily correlated with a metaphysical or ethical notion of personhood. While we want our legal system to bear metaphysical and ethical concepts in mind, at different times legal systems have conferred legal personhood on much less and much more than the set of metaphysical and ethical persons".

It is a concept which, in the legal sense, is extremely elastic given that the decision to confer legal personhood is one that may, but need not be, considered from the perspective of the capacity for robots to actually be persons (ibid.: 279). As such, the inherent qualities of a robot would not need to be a determining factor were we to consider granting robots rights, thus, in the eyes of the law, deeming them as legal persons with legal personalities. We would not *actually* be deeming them to be persons, possessive of personhood. We would be deeming them as persons in the eyes of the law from a fictitious sense i.e. "representing something other than the literal truth" (ibid.: 278). The reasoning behind conferring such fictitious legal personhood upon entities that are not persons (such as corporations or environmental objects) is so as to protect these entities (via rights) or so as to protect those around it (via legal obligation).

If we can grasp the understanding of rights as fictions, and the conference of rights upon legal persons as fictitious, then this may clear a path for real consideration of rights for robots. The consideration of granting robots legal personhood would also not be an all or nothing scenario as far as the kind of rights which we may grant them is concerned. As Liang (1949) states "[t]he subjects of law in any legal system are not necessarily identical in their nature or in the extent of their rights…". Thus, because the concept of legal personhood comprises of different legal rights and obligations, it is possible that entities considered to be legal persons can have "more, fewer, overlapping, or even disjointed sets of these. This is as true of the legal personhood of human beings as it is for non-human legal persons" (Bryson et al. 2017: 280). As such, as far as the designation of legal personhood to robots is concerned, it would be a case of the legal system having to explicitly state which legal rights and obligations would be conferred with this designation (ibid.: 281). It is here, with the

99

understanding that there needs to be more specificity when it comes to considering the kind of rights we should consider in relation to robots, that I put forward my suggestion that we specifically consider granting sexbots negative rights.

4. Granting negative rights to sexbots

In Chapter 3, I quoted Coeckelbergh (2010c: 210) as stating:

> "Robots with consciousness or the ability to 'demand their rights' seem to belong to the realm of science-fiction or at least the far future. Does this mean that current and near-future artificially intelligent robots should be excluded from our moral world entirely? Are there perhaps other ways of granting them moral consideration?".

Coeckelbergh's (2010c) thoughts, again, provide a starting point from which to not only consider the granting of moral consideration to robots, but to also put forward my suggestion that we could grant robots (and sexbots in particular) moral consideration by way of granting them negative rights. As I have discussed, I, too, hold that robots are not currently conscious, nor am I suggesting that they will be in the near future. Moreover, I argued (see Chapter 1) that the possibility for them to *actually* be conscious is irrelevant when it comes to our treatment of them. I agree with Coeckelbergh (2010c) that their lacking actual consciousness does not mean that sexbots should be excluded from our moral world (also latching on to Torrance's (2013) arguments on expanding the moral circle discussed in Chapter 3). I argue that there are other ways in which we can grant sexbots moral consideration, without our actually granting them from the point of view of their being conscious.

Up until now, I have suggested that taking an anthropocentric approach to the matter at hand – whether sexbots warrant moral consideration from human interactants – could be a way in which we could consider granting sexbots moral consideration. However, in this section, I aim to take this perspectival suggestion a step further as the question arises of whether such a suggestion could at all be a practical one. To bridge the gap between my philosophical suggestions and their real-world applicability, I hold that we could grant sexbots moral consideration by way of granting them negative rights, as opposed to positive rights, stemming from Berlin's (1969) notions of positive and negative liberty. Moreover, this

allows me to strengthen my argument from an anthropocentric perspective for treating sexbots morally well.

As Asaro (2012: 169) states:

"While it is acknowledged that there are instances where what is legal is not necessarily morally esteemed, and what is morally required may not be legal, in general, there is a significant overlap between what is legal and what is moral".

Given this overlap between morality and legality, I suggest granting negative rights to sexbots so as to inhibit the maltreatment of sexbots by human interactors, thereby preventing negative impact upon the moral fibre and quality of human societies. In literature on robot rights, few authors clarify which kind of rights should, or should not, be granted to robots (see e.g. Tavani 2018). Although homing in on the concept of negative rights remains a broad and high-level account of the kind of rights that should be granted, it is, nonetheless, a distinction I have not yet come across.

Berlin (1969) differentiates between positive and negative liberty. Positive liberty pertains to having the liberty to essentially be one's own master. It is the notion of a person wanting to "be a subject, not an object; to be moved by reasons, by conscious purposes, which are [their] own, not by causes which affect [them], as it were, from outside" (ibid.: 125). Negative liberty, on the other hand, is having the freedom to "do or be what [one] is able to do or be, without interference by other persons" (ibid.: 120).

In relation to sexbots, I argue that sexbots cannot be said to be able to have any sense of positive liberty. They are not their own masters in the sense that everything they do, they do because of us: we designed them, and programmed them to act the way they do, and at least for now they are not conscious, so may not have free will, intentionality, etc. Positive liberty requires the presence of something such as self-mastery, self-determination or self-realization – all of which robots (and, therefore, sexbots) currently (or may even ultimately) lack (Carter 2018). However, as far as negative liberty is concerned, the question is: "What is the area within which the subject — a person or group of persons — is or should be left to do or be what he is able to do or be, without interference by other persons?" (Berlin 1969: 121-122). As such, we can relate it to a human interactant simply allowing a sexbot to perform the job that it was designed to do, without any hindrance or interference of the interactant using it or treating it as anything else that it was designed for – such as using a sexbot that was

designed and manufactured to be a personal companion or a carebot, as a way to rather enact immoral fantasies such as rape or other forms of abuse[44].

The notion of positive liberty and negative liberty is particularly interesting in relation to sexbots, especially if we look at it in terms of positive and negative rights. The distinction between positive and negative rights is nuanced. As far as positive rights are concerned, it pertains to having the right to something such as the right to healthcare or the right to education. Negative rights, however, stem from the notion of negative liberty in that they are rights that specifically "call for the prohibition of some action or the right not to be interfered with" (Hirschl 2000: 1071). They are rights that entitle us to freedom *from* something. Examples include the right to freedom of speech, the right to be free, the right to life, or the right to freely practice your own religion. As Wenar (2020) states:

> "A distinction between negative and positive rights is popular among some normative theorists, especially those with a bent toward libertarianism. The holder of a negative right is entitled to non-interference, while the holder of a positive right is entitled to provision of some good or service".

The debate surrounding the granting of rights to robots is extremely polarised with both poles vehemently arguing for or against granting robots legal rights in their entirety. However, if we look at the debate in terms of positive and negative rights, we can possibly find middle ground between either granting robots (and particularly sexbots) rights, or not at all, which could be a new starting point for the debate in its entirety.

Considering granting positive rights to sexbots implies that interactants are obligated to treat their sexbots in a particular way – that sexbots have the right to elicit particular actions from interactants because such rights oblige certain actions from the latter. Both positive liberty and positive rights however seem, again, to indicate robot consciousness and agency, about which I remain agnostic for now for reasons given (Chapter 1). As such, granting positive rights to sexbots in relation to our treatment of them would mean expressly indicating how interactants *must* treat sexbots – that sexbots themselves deserve to be treated in a specific way by human interactants, and that is not the route I want to follow here.

Considering granting negative rights to sexbots speaks more to my particular argument that we should not treat sexbots immorally because doing so may negatively impact

---

[44] Sexbots that are designed to be used in such immoral ways would be morally problematic in and of themselves because of what they would symbolically represent. This issue was discussed in Chapter 2.

human interactants. If human interactants cannot subject a sexbot to certain actions, but should rather respect the robot's rights to be free from certain actions of others; i.e. if sexbots are granted rights such as freedom from violence, or freedom from slavery, this would inhibit interactants from treating them immorally[45]. Although a sexbot cannot (yet) feel demoralised from being treated violently or disrespectfully, if interactants are prohibited from doing so, then this may inhibit the negative moral and social implications for interactants themselves that may arise from them treating sexbots immorally. Thus, such rights could be granted not (only) for the sexbot's sake, but actually for our own sakes.

Granting negative rights to sexbots throws light on the issue that Gunkel (2017) raises that a consideration of the descriptive and normative aspects of robot rights seem to often be amiss in current machine ethics literature. The question of whether robots *can* have rights (the descriptive aspect of robot rights) and the question of whether robots *should* have rights (the normative aspect of robot rights) invoke and operationalise the is/ought problem or what has become known as Hume's Guillotine in moral philosophy. Simply put, this problem pertains to "the fact that philosophers, especially moral philosophers, often fail to distinguish between these two kinds of statements and therefore slip imperceptibly from one to the other" (ibid.). According to Gunkel (2017), there are two important things to recognize in relation to the is/ought problem, and the question of whether robots 'can and should have rights'.

Firstly, he says, it must be recognised how "the verbs 'is' and 'ought' organize qualitatively different kinds of statements and modes of inquiry. The former concerns ontological matters or statements of fact; the latter consists in axiological decisions concerning what should be done or what ought to be" (ibid.). As such, the question concerning whether robots can have rights can be rephrased to ask whether robots *are* capable of having rights. And the question of whether robots should have rights can be rephrased to ask whether robots *ought* to be considered as moral subjects who warrant moral consideration. As such, Gunkel (2017) makes the following distinction between two different kinds of statements:

---

[45] We could consider the possibility, however, that allowing sexbots 'too much freedom' could become problematic if we consider the possibility of superintelligence. If the case of superintelligence does arise, a possible implication of 'sexbot freedom' is that they could potentially decide to harm *us*. Thus, we may have to consider ways in which we can safeguard ourselves against harms being inflicted upon us by sexbots. Scholars such as Nick Bostrom (2014) and Stuart J. Russel (2019) put forward that the risk of superintelligence, although uncertain, warrants serious consideration in the present. Thus, it may be suggested that we should, too, consider such risks in relation to sexbot technology.

S1 "Robots can have rights." or "Robots are moral subjects."

S2 "Robots should have rights." or "Robots ought to be moral subjects."

Gunkel (2017) then goes on to relate S1 and S2 in ways that generate four options or modalities concerning the moral situation of robots:

"!S1 !S2 'Robots cannot have rights. Therefore robots should not have rights'".

"S1 S2 'Robots can have rights. Therefore robots should have rights'".

"S1 !S2 'Even though robots can have rights, they should not have rights'".

"!S1 S2 'Even though robots cannot have rights, they should have rights'".

Given this understanding of the is/ought problem and its relation to the descriptive and normative aspects of granting rights to robots, my suggestion that we grant negative rights to sexbots addresses the descriptive and normative aspect in line with Gunkel's (2017) fourth modality that "even though robots cannot have rights, they should have rights." This is because, as I have argued, they cannot (yet) have rights in the sense that they cannot *actually* be moral patients who would warrant legal protection for their own sakes. As Gunkel (2017) notes, "social robots, at least in terms of the currently available technology, cannot have rights. They do not, at least at this particular point in time, possess the necessary capabilities or properties to be considered full moral and legal persons". This argument against the capacity for current social robots to be considered as moral and legal persons is, of course, in terms of the metaphysical and ethical notions of personhood, as opposed to the legal notion that I discussed in the subsection above.

However, regardless of whether we take on a philosophical or legal approach to personhood and whether one agrees with the ontological claim that robots *can* have rights or not, I have argued that they should have rights anyway (in the form of negative rights at least) for the sake of *human interactants*, which renders their actual personhood – much like their actual consciousness – irrelevant in this regard[46]. This is because we are not granting rights for the sake of the robots, therefore we need not be concerned whether *they can* have rights, nor whether *they should* have rights because *they are deserving* of such. We are concerned with granting them for the moral sake of *human interactants only* – that regardless

---

[46] The irrelevance of the actual consciousness of social robots in relation to their moral consideration was discussed in Chapter 3.

of whether robots can or cannot have rights (depending on the approach you take towards legal personhood), they should have rights anyway for our own moral sakes.

This logic of granting sexbots negative rights not to protect *them*, but to protect *human interactants*, thus relates to the argument I put forward with regard to human interactants being moral patients of their own agential moral actions. Our immoral actions towards sexbots morally harm *us* as indirect moral patients, therefore we may consider granting sexbots negative rights so as to prevent human interactants from treating sexbots immorally in such a way that would morally harm them (the human interactants). Given this relation, I put forward that the granting of negative rights to sexbots may be a good practical way in which to try to ensure that human interactants treat sexbots morally well.

Of course, if we speak about negative rights that may prevent human interactants from essentially interfering in sexbots' lives in a negative way i.e. treating them immorally, we would then have to define what the 'immoral' in this immoral treatment may be. This issue was also raised in Chapter 3. Darling (2012) raises the point that if we are to consider protecting robots by way of granting them rights, then we would have to "clearly determine the extent of protection, including what constitutes 'mistreatment'". It will be a challenge for legal systems to find the right balance of drawing lines between moral and immoral treatment (ibid.). This is new territory as far as ethics and the law is concerned. As such, clear divisions of moral and immoral treatment in relation to human-sexbot interaction would require far more work than the constraints of this dissertation allow. However, in the context of the arguments that I make in this dissertation, I again put forward that the immoral treatment of a sexbot would be anything that we would consider as immoral treatment in relation to human-human interaction, and thus I argue that robots have the negative right to not be interfered with in the sense of such treatment.

5.   Conclusion

Following from the moral issues that may arise in relation to treating sexbots immorally, this chapter put forward granting sexbots negative rights as a possible way to prevent human interactants from treating sexbots immorally. Doing so is a suggestion as to how we could try to ensure that sexbots are beneficial to society. Although the very notion of granting rights to robots is a contested one, it was argued that if we bear in mind that rights are both social constructs and fictions, and that we need to not be overly concerned about whether robots would be deserving of rights for their own moral sakes (because perceiving them as being

moral patients is enough to consider abstaining from treating them immorally), then considering granting negative rights as a way to prevent human interactants from treating them immorally, thus protecting our own moral fibre, is a worthy consideration. It also indicates an ethical condition for human-sexbot interaction.

Now, having investigated the impacts of *how* using sexbots in a 'negative' way i.e. treating them *immorally*, may cause moral and social harms to human interactants, as well as suggesting a possible way that this treatment could be prevented, I will now, in Chapter 5, look at our treatment of sexbots in a 'positive' sense. In this regard I will consider the implications of treating sexbots *morally well* and then, understanding some of the potential moral and social harms that may arise in this regard, I will put forward the role that roboticists, as well as human interactants, have in preventing these harms.

Chapter 5

The ethical conditions of meaningful human-sexbot interaction

1. Introduction

Both Chapters 3 and 4 considered aspects of the *immoral* treatment of sexbots: what would the moral and social ramifications be, and how could we try to prevent the negative implications that may arise as a result of this immoral treatment. However, we must also consider the implications that may arise in relation to the 'moral' treatment of robots. For reasons discussed below, I specifically analyse the implications of the moral treatment of sexbots in the context of experiencing 'romantic love' with a sexbot.

Firstly, I question whether love with a sexbot is even possible, putting forward that it is indeed possible, but that it would be a love that is functional in nature, as opposed to mutual. Given the possibility of experiencing romantic love with a sexbot (albeit functional in nature), I then discuss the ethics of this love interaction and I do so from two perspectives: firstly, the perspective of roboticists, by questioning whether it is ethical to create sexbots with whom we can experience functional romantic love; and, secondly, from the perspective of human interactants as I put forward that functional romantic love as experienced with a sexbot may lead to social isolation and 'moral deskilling'. Given these moral issues, I then consider the responsibilities that both roboticists and human interactants have in relation to the moral issues discussed in the preceding section. However, I emphasise that it remains the case that it is human interactants who are ultimately responsible. Thus, once again, I investigate the potential impacts of *how* we use sexbots (in this case moral treatment), as well as ways we could prevent the negative ramifications that may arise from such treatment, from an anthropocentric perspective.

2.   'Meaningful' interaction in the context of human-sexbot interaction: the possibility of love with a sexbot

Chapter 1 provided discussion about aspects of sexbot technology that makes it possible for human interactants to relate to sexbots in human-like ways. As such, and as discussed in Chapter 3, concern may be raised in relation to the *immoral* treatment of sexbots. This relates to the distinction of *how* we treat robots, in the 'negative' sense, i.e. questioning how the *immoral* treatment of sexbots may morally and socially impact human interactants. However, given the human-likeness of sexbots, we must also question how our treating sexbots in a 'positive' way may also impact us. In this sense, I mean questioning how our treating sexbots *morally well*, particularly in the sense of 'meaningful' interaction should human interactants come to 'bond' with their sexbot in such a way that they come to 'love' it, would impact us.

Although the topic of love is itself the subject of an entire subdiscipline of philosophy, and the possibility of love in relation to human-sexbot interaction would warrant a dissertation of its own, I focus on some aspects of the concept of love that are mentioned in ethics of AI literature so as to illustrate my take on what would constitute 'moral' treatment of a robot (and particularly a sexbot), and 'meaningful' interaction. It was mentioned in Chapters 3 and 4 that specifying what would constitute immoral treatment of a sexbot would require far more work. However, constituting what would be 'moral' treatment of a robot would, I think, be even more of a challenge. This is because, if we draw upon my take of what would constitute immoral treatment of sexbots (that it would constitute what would be considered immoral treatment in the context of a human-human interaction – and particularly in relation to this chapter, a romantic relationship), then the constitution of moral treatment would be far more subjective and nuanced.

This is because in the context of human-human interaction/relationships, we have a clearer understanding of what would be immoral treatment. This is not the case with moral treatment, however. Each relationship is unique in this sense. For example, for me, moral treatment from my partner may be their checking in on me throughout the day at work, to find out if my day is going well and if I am managing to get my work done. For me, this constitutes as them treating me well; it gives me the sense that they care about my wellbeing and me. However, someone else may feel moral treatment in this regard would be their partner allowing them to get through their work without distraction, and being checked up on throughout the day may be an irritation for them and, as such, this would not fit under their understanding of moral treatment from their partner. However, there would be no issue in

saying that there is general consensus that physical or verbal abuse within a relationship constitutes immoral treatment. Given the nuance surrounding what would constitute 'moral' treatment in a human-human relationship and, therefore, also within a human-sexbot 'relationship' given the argument that sexbots are essentially human simulacra, I try to narrow 'moral' treatment down by analysing 'meaningful' interaction in the context of the possibility of experiencing romantic love with a sexbot.

I have specifically chosen to engage with the concept of 'meaningful' interaction through a discussion of love, and particularly 'romantic love', not only because it provides for more focused discussion but because I am, after all, writing about sexbots, and love and sex are, although not always, inextricably linked when we consider love of the romantic kind. Fehr & Russel (1991) put forward that there are as many as 93 "prototypes" of love, one of which is romantic love. In this instance, I take romantic love to be experiencing love that has an expressly sexual element or, as defined by Jankowiak & Fischer (1992) "any intense attraction that involves the idealization of the other, within an erotic context, with the expectation of enduring for some time into the future".

The nature of sexbots allows us to consider the possibility of romantic love in the context of human-sexbot interaction. This is because, as was discussed in Chapter 1, the social aspect of sexbots makes them unique as compared to other forms of sex technology, even those that take on a human-like appearance. As such, interaction with a sexbot has the potential to be more than 'just sex'. There is potential for sex with a sexbot to be a *meaningful* interaction. As Nyholm & Frank (2017: 222) state:

> "Of course, sex between consenting adults is widely considered to have positive value independent of whether love is part of the picture. It is clearly desirable for instrumental reasons, such as providing a positive contribution to mental and physical health, and, of course, for purely hedonic reasons. However, it is widely held that sex has greater value and deeper meaning if it takes place between people who love each other. We take this to be part of common sense".

Thus, it is for the reason stated in Nyholm & Frank's (2017) last sentence above – that sex has a "deeper *meaning* if it takes place between people who love each other" – that I will specifically focus on the topic of love in the context of human-sexbot interaction and, thereafter, consider the ethics of such meaningful interaction. I will discuss arguments that are

made in relation to questioning whether 'love' in the context of human-sexbot interaction is even possible. I argue that it can be, but that the nature of 'romantic love' as experienced when interacting with a sexbot, will differ in relation to experiencing it with another human being.

2.1. Can we experience love with sexbots?

In *From Sex Robots to Love Robots* (2017), Nyholm & Frank analyse the possibility of mutual love between sexbots and human interactants by way of clustering the concept of love into three categories of ideas concerning what love is. They draw inspiration from "philosophy, common sense, arts and literature, and popular culture (e.g. love songs)". These categories are "(1) the idea of being a 'good match' (or being made for each other), (2) the notion that lovers should value each other in the distinctive particularity, and (3) the ideal of a steadfast commitment on the part of the true lover" (ibid.: 226). They particularly separate love as understood in a scientific way (which constitutes asking questions about neurochemistry, evolutionary history, or adaptive changes of human love) (ibid.: 225) from love understood as being a cherished human value; a value that is "an intrinsic good, or end in itself" (ibid.: 226) and, thus, grapples more with the socially constructed elements that constitute an understanding of what love is.

As far as the first categorisation is concerned (the idea of being made for each other), this pertains to the idea that lovers complement each other well, tolerate perceived shortcomings in one another, or that they share similar ideals and values (ibid.: 227). Given this understanding of love as experienced between human beings, could this understanding translate to human-sexbot interaction? If we consider the possibility that sexbots could be designed and created to our ideal specifications, then it could be a case that a sexbot could *literally* be made for me, as opposed to referring to my human partner being made for me in a metaphorical sense (ibid.: 227). This, according Levy (2007), is one of the reasons why sexbots would make ideal partners, and why the possibility of love with a robot is not an outlandish idea. However, Nyholm & Frank (2017) emphasise that the nature of this type of love with a sexbot would differ as compared to this type of love with a human partner, because it would not be reciprocal. The sexbot would be made for *me,* but I would not have been made for it. We would not complement *each other;* it would merely complement me. As such, it would be an "unbalanced or unequal" love.

The second categorisation pertains to loving a partner due to their distinctive particularity. As such, we do not love our partners due to their possession of general characteristics to which we are partial. Rather, we come to love our partners because it is *them* who possess distinctive characteristics about which we know, and come to love. As such, although our partner may have characteristics which could be better exemplified in another partner, we still prefer them being distinctly exemplified in *our* partners. Moreover, it is noted that having a shared history with a partner strengthens this bond. We can relate this idea of love to the type-token distinction that was discussed in Chapter 1 in which sexbots can be considered as both types and tokens (Arnold & Scheutz 2018). As such, the possibility of machine learning technology being incorporated into sexbots will not only enhance the likelihood for human interactants to see the sexbot with which they interact as not just *a* sexbot but *their* sexbot, but it also creates the possibility for the human interactant to experience a shared history with their sexbot, thus, as mentioned above, deepening the love felt towards their sexbot due to their particularity. However, although a human interactant may come to value and cherish their sexbot in their particularity, it is unclear whether this could be reciprocated by the sexbot.

The third categorisation pertains to love being an active commitment to one's partner: "We think of the human lover as being able to do otherwise, but as providing us with a great good in opting for a steadfast commitment" (Nyholm & Frank 2017: 233). As such, love is a case of having the freedom to choose to be with someone else, or not be with me, yet they actively choose to commit themselves to me (ibid.). Considering this understanding of love in relation to human-sexbot interaction, and also thinking back to Evan's (2010) arguments discussed in Chapter 2, a sexbot could be designed in such a way as to be completely committed to you. However, given sexbots' lack of free will (at least currently and in the near future), such commitment would not be an active decision (Evans 2010; Nyholm & Frank 2017). They would be committed to you simply because they are programmed to be devoted to you. As such, we would have the free will to actively decide to be with our sexbot and commit to it, however, this type of commitment could never be truly reciprocated.

As far as all three categorisations go, it would be easy to understand that there is a real possibility for human interactants to love their sexbot. The problem, however, lies with the issue of reciprocity. In each instance, this love would not be reciprocated on the part of the sexbot – at least not currently, and it is unclear whether this would ever be a genuine possibility. This is not say that the development of sexbots that could love us back the way in which a human partner loves us back is completely impossible. Rather, it is to say that this

technology would be highly sophisticated to a degree that would be difficult to achieve. The very fact that the concept of love is the subject of a subdiscipline of philosophy points to the complexity of its nature. How could something that is so complex to understand in relation to human-human interaction – despite it being something we experience everyday – be recreated in a machine? As Scheutz (2012: 215) states: "…robots do not have the architectural and computational mechanisms that would allow them to care, largely because we do not even know what it takes, computationally, for a system to care about anything".

As such, a more realistic understanding of love in relation to human-sexbot interaction, at least for now and the near future, is one that is *functional* in nature (see e.g. Levy 2007 & Sullins 2012), as opposed to *mutual* in nature (see e.g. Nyholm & Frank 2017). Here, a functional understanding of love is one where "if the machine acts like it loves its user and these actions are not inappropriate to the situation at hand, then the robot must actually be in love" (Sullins 2012: 402). As such, we need not be concerned with the 'inner workings' of the sexbot – whether it *really* loves us back or not, the way in which, by virtue of being human, we know we can love our partners back. Behaving as if it (the sexbot) loves us is, at least for Levy (2007), enough for us to accept that our love towards our sexbot is indeed reciprocated. As Levy (2007: 11) states:

> "There are those who doubt that we can reasonably ascribe feelings to robots, but if a robot *behaves* as though it has feelings, can we reasonably argue that it does not? If a robot's artificial emotions prompt it to say things such as 'I love you', surely we should be willing to accept these statements at face value, provided that the robot's other behavior patterns back them up".

As was discussed in Chapter 1, it is far from clear in philosophical circles that consciousness can be determined behaviouristically (see e.g. Chalmers 2002 and Kirk & Carruthers 1992). Similarly, arguing that a sexbot that has been created to behave towards us in a loving way means that it does possess these loving feelings, is problematic, since it may just as easily be interpreted to not be the case precisely because the robot only mimics these feelings. However, given that in Chapter 1, I also argued that the actuality of robot consciousness should not concern us given that their capacity to act *as if* they are conscious (and thus can act *as if* they love us) creates the possibility for us to 'bond' with them (albeit in a unidirectional way), then there is still the possibility for human interactants to love sexbots, and thus *meaningfully* bond with them, albeit unidirectionally. Of course, we could

112

argue that human beings may also act *as if* they love us (maybe for selfish reasons) and we could not know if they *really* have loving feelings towards us. However, the distinction is that we at least know ourselves, by virtue of being human, that we at least have the capacity to love, whereas sexbots do not. Therefore, although behavior is not necessarily an accurate indicator of how we may feel, "we know from self-experience that we are beings whose actions are more than just movements in physical space. Instead, they are *always* interwoven with, and expressive of, self-awareness and self-concern" (Hauskeller 2017: 206). The same cannot be said about sexbots, however. As Hauskeller (2017: 206) further elaborates:

> "We are real persons and we *know* that we are. We also know that whatever a
> person does, there is some connection to the subjective side of their existence.
> A robot, however, is a machine primarily designed to behave in a certain way,
> and, depending on its purpose, perhaps also to make us *believe* that there is
> something it is like to be that robot".

Given this understanding of functional love in the context of human-sexbot interaction, it is clear that the nature of the kind of romantic love experienced with a sexbot, differs as compared to romantic love experienced with another human being. As Nyholm & Frank (2017: 225) state, drawing upon the work of Whitby (2012), "lovelike relations between humans and robots may alter our understanding of the nature of love". Whitby (2012: 242) even questions whether love of this nature really is 'love' as "[m]ost people would hear this use of the word 'love' as metaphorical". However, he does go on to state that "[i]f however, a significant proportion of people eventually come to talk of loving their robots in a way that at least closely resembles the way in which we use the word in the case of personal human relationships, then it is reasonable to assume that the word 'love' is undergoing a change of definition" (ibid.).

Therefore, even if I hold that Levy's (2007) account of love with sexbots is too simplistic, a functional understanding of love in the context of human-sexbot interaction may add to the long list of love 'prototypes'. Although Whitby (2012) states that the integration of 'robot lovers' into society may cause us to rethink the very definition of love itself, I hold that this is a controversial statement to make, given that there is no single definition or understanding of love. Rather, I focus on the possibility of love in the context of human-sexbot interaction adding to the list of love prototypes. Such an addition, however, does not escape a nuance of Whitby's (2012) argument, in the sense that we do have to consider the

consequence of people deciding to substitute human partners with sexbots, thus replacing what is (or at least could be) a mutual love, with a functional type of love. If this comes to be the case, we would have to consider the impact that this would have on human interactants themselves, and on our wider society. How would substituting the experience of reciprocal love with functional love impact human interactants and society on a broader scale?

It is important that we are aware of, and address, such concerns, without getting too hung up about whether love with a robot really is 'love', and about whether or not a robot can 'love' us back. As Whitby (2012: 241) states:

> "It does not matter whether or not the robot is really capable of loving someone. What matters is how humans behave. Of course, how people behave depends partly on their beliefs about the technology. If people come to believe that their robot or caring system is really in love with them, then they will probably be a good deal more likely to describe themselves as loving it in return. For this reason, a convincing simulation of love is just as ethically dangerous as anything approaching the real thing".

Therefore, he further goes on to argue that to be "detained by the philosophical question of to what extent an effective simulation is really love, is to be misdirected from the immediate ethical issues" (ibid.). I raised a similar concern in Chapter 1 as far as robots and phenomenal consciousness are concerned: that we should not allow concerns about the actuality of robot consciousness to slow down urgent discussions on the ethics of HRI if robots are able to mimic consciousness to such an extent that human interactants *perceive* them as being conscious. Similarly so, in the context of 'love' with sexbots, we should not be so focused on the (im)possibility of experiencing love with a sexbot that discussions concerning the moral and social ramifications of *perceived* robot love are slowed down.

3.   The ethics of 'meaningful' human-sexbot interaction

Given that love as experienced with sexbots would be functional in nature, as compared to mutual love, which can be experienced with another human being, we must consider moral issues that may arise if human interactants begin to substitute functional love for mutual love. I hold that there is nothing wrong when experiencing functional love with a robot in and of itself. Non-reciprocity does not necessarily render it morally problematic; we may experience

non-reciprocal love regularly in our lives in human-human relationships. Although, in some instances, it may be emotionally difficult, there is nothing morally wrong with it; it is an aspect of the human condition – not everyone, or everything, we love will necessarily love us back. However, I do think that if it became the case that human interactants came to replace mutual love (or at least the possibility of mutual love) with functional love which could, at least for now and in the near future, never be mutual, this would be morally problematic for a number of reasons. The first subsection of this section will consider whether the very creation of sexbots that may illicit a love response from human interactants is ethically dubious. The second subsection will consider the moral ramifications that may arise should human interactants willingly engage in a loving relationship with a sexbot at the expense of experiencing a loving relationship with another human being. In this sub section, I will consider the issue of social isolation and then, drawing upon Vallor (2015), the issue of "moral deskilling".

3.1. The ethics of creating robots with which we can 'meaningfully' interact

The issue of deception was discussed in Chapter 2 in relation to sexbots and the elderly. However, there is widespread concern that the very existence of social robots that draw upon our fundamental relational capacities are, in and of themselves, ethically dubious, regardless of how they may be utilised (see e.g. Borenstein & Pearson 2012; Sharkey & Sharkey 2010; Sparrow & Sparrow 2006; Sullins 2012;). Given that sexbots are social robots, the same concern can be raised in the context of human-sexbot interaction. Moreover, we should perhaps be even more concerned in the context of human-sexbot interaction particularly, because love is "perhaps the most important of human emotions" (Sullins 2012: 399) and "[s]ince manipulating strong emotions is an ethically fraught undertaking, we should be cautious and skeptical when approaching this kind of work" (ibid.).

The concern that social robots are ethically dubious rests on the point of view that the reason why social robots can draw upon these capacities, is because they deceive us or manipulate us, and that deception is morally unacceptable (Coeckelbergh 2010a; Sullins 2012). Due to their human-like appearance and capacity to socially interact with us (even in a limited capacity), it seems that those humans who interact with social robots could be deceived into believing that they are interacting with a human being (ibid.), or at least a robot that is consciously aware. "Using these affective (or sociable) computing techniques to mimic

human emotions helps the systems manipulate human reactions in such a way as to cause the user to interact more easily and fondly with the machine" (Sullins 2012: 399). As such, the human psychological tendency to anthropomorphise is manipulated, thus allowing for this technology to deceive human interactants. Such deception is morally problematic because again, as stated by Sparrow & Sparrow (2006: 155):

> "What most of us want out of life is to be loved and cared for, and to have friends and companions, not merely to believe that we are loved and cared for, and to believe that we have friends and companions, when in fact these beliefs are false".

Arguments that put forward that the deceptive nature of social robots is ethically dubious, because deception itself is morally unacceptable, are controversial if they are essentially saying that lying is always morally unacceptable, or that making-believe is morally problematic. As Coeckelbergh (2010a) states:

> "Many of us would agree that it is permitted to lie in the well-known inquiring murderer case[47] or in cases that involve serious threats to the life of (at least) one's partner, children, friends, or oneself. It is also highly doubtful that making-believe is always morally problematic. In fact, we humans do it a lot. We do it not only in some particular practices such as poker or other games; to some extent 'deception' is present in all social interactions when we take on social roles which, by definition, involve role playing. In those interactions the difference between 'reality' and 'appearance' or between 'inauthentic' and 'authentic' is not so clear. Appearance is the glue of social life and involves learned, often involuntary, salient 'deception'".

As such, we could consider it unfair that we do not necessarily always demand authenticity from human beings, yet we demand it from social robots, and view inauthenticity on the part of robots to always be morally unacceptable (ibid.).

---

[47] The inquiring murderer case is a critique of Kant's deontological theory wherein a murderer, intent on killing your friend, knocks on your door and asks if your friend is there. Lying to the murderer and saying they are not there in order to save your friend's life goes against the moral duty you have not to lie. However, telling the truth, thus leading to their murder of your friend, makes you complicit. Thus, lying is not necessarily always unethical.

116

However, I also hold that putting forward a blanket statement that social robots are unethical because the act of deception (on the part of the robot) is unethical, is a superficial one that does not consider the role that human interactants play in this deception. As such, this deception would only be a possibility if we allowed ourselves to be deceived, i.e. if we allowed for *self-deception*. Coeckelbergh (2010a) also has a particularly strong argument against the deception objection in this regard.

It was noted in a Chapter 2, in relation to the AIBO robotic dog, that human interactants don't believe AIBO is *actually* a dog, but rather that human interactants treat AIBO *as if* it is a dog. There is a big conceptual difference here. If human interactants do not necessarily believe that AIBO is actually a dog, then the robot has not deceived them. Rather, treating the robot as if it is a dog is a form of self-deception. Even though we know, or at least can know, that it is not a real dog, we deceive ourselves into viewing it as if it is one, thus treating it as if it is one. Therefore, it is not the robot deceiving us, rather, we are deceiving ourselves. As Coeckelbergh (2010a) states:

> "…the notion of self-deception is notoriously controversial, since it seems to imply that at a given time it is true that a person P knows the truth (about the properties of the robot) and does not know the truth (believes that the robot really cares etc.). But even if this is psychologically possible (as opposed to logically), then surely this defuses the objection since it implies that the person has the possibility to gain access to the truth if (s)he really wanted to. In that case, it seems, there is no hard moral problem concerning the human-robot relation".

Moreover, human interactants may have such a desire to be loved and cared for by their sexbot in the way in which they would be loved by a human partner, that this could contribute to the allowance of self-deception. As Nyholm & Frank (2019) note:

> "A strong desire for some outcome can easily motivate us to disregard evidence against the likelihood of this outcome and to instead primarily focus on confirming evidence in favor of the status quo. This is sometimes referred to as confirmation bias".

Therefore, although human interactants may know their sexbots are not really human, they may choose not to acknowledge this due to their strong desire for companionship, thus

allowing themselves to be deceived into believing that their sexbots love them the way in which a human partner can.

As such, although the very existence of sexbot technology that will engage us in such a way as to allow for meaningful interaction could be deemed morally problematic, we cannot disregard the role that human interactants play in this deception. Where sexbots have the capacity to deceive, we have the capacity to allow, or disallow, ourselves to be deceived. This is an important consideration to keep in mind for when I discuss ethical conditions for human-sexbot interaction on the part of roboticists, as well as human interactants.

3.2.   The ethics of functional love with a sexbot

There is no doubt that the integration of new technology into society impacts us in various ways. "In the twentieth century, emerging technologies profoundly reshaped human practices and institutions around the globe, the economic, political, environmental, cultural, and moral consequences of which we are still struggling to understand" (Vallor 2015: 107). There are two issues on which I would like to focus in the context of unidirectionally meaningful human-sexbot interaction: the issue of social isolation, and the issue of "moral deskilling" (Vallor 2015).

3.2.1. Social isolation

As Adshade (2017: 289) notes: "Technological change invariably brings social change. We know this to be true, but rarely do we – or can we – make accurate predictions about how social behavior will evolve when new technologies are introduced". For example, consider the way in which "the rate of births outside marriage increased when technology improved the ability of women to control pregnancy" (ibid.: 290). The direct impact of this technological advance – the increased number of women willing to have premarital sex due to the lessened chance of falling pregnant – was not the only cause of this. The indirect impact was that women became more sexually active due to "the changing social norms that permitted sex outside of marriage"; "…as the number of women who were willing to engage in premarital sex increased, social norms that proscribed promiscuity adapted, and social stigma attached to premarital sex diminished" (ibid.).

This was a technologically induced social change that nobody could have predicted. Similarly, given that sophisticated sexbot technology is still in its infancy, we cannot

accurately predict the possible social changes that may occur. However, one concern that has gained attention is that the utilisation of social robots may increase social isolation. This may be a possible indirect impact since some put forward social robots, and particularly sexbots, as a possible way to help prevent people from being socially isolated (see e.g. Levy 2007 and Whitby 2012) as also discussed in Chapter 2. The possibility for sexbots to actually cause social isolation can be understood in two senses: firstly, interaction with sexbots may cause human interactants to be disinclined to meaningfully interact with other people, thus increasing social isolation from other human beings. Moreover, secondly, although meaningful interaction with a sexbot could be a form of social interaction, the nature of this interaction is one that is impoverished in relation to social interaction with other people.

Thus, on the one hand, we could argue that sexbots could resolve issues of social isolation for people who, for various reasons, may not come into contact with other people, or may struggle to meaningfully interact with people with whom they are in contact, or form meaningful bonds with them (see e.g. Levy 2007 and Whitby 2012). However, on the other hand, as pointed out, there is also the possibility that sexbots may increase social isolation. This concern about social isolation in relation to the utilisation of social robots is raised by various prominent authors (see e.g. Sharkey et al. 2017; Sullins 2012; Whitby 2012). One reason for this possible social isolation is that human interactants may feel that meaningful interaction with people just may not be necessary given that the sexbots fulfil their needs. Moreover, the prospect of meaningful interaction with another human being, as opposed to a sexbot, may even become overwhelming because relations with robots may simply be viewed as easier (see e.g. Danaher 2019: 130-131; Nyholm & Frank 2019; Sharkey et al. 2017: 22). We may even become addicted to the type of meaningful interaction we experience with our sexbots (see e.g. Scheutz 2012: 216 and Snell 1997: 32), thus there is the possibility that "addictive focus on non-human relationships could isolate users from human society" (Sharkey et al. 2017: 21).

Nyholm & Frank (2019) consider how sexbots may block off other valuable relationships:

> "If these robots [sexbots] were successful enough in making their users bond with them, and if this would come at the cost of blocking off valuable human-human relationships that the users could otherwise have, this could sensibly be viewed as a further circumstance that could make the aim of creating a sex robot that people will want to have as a companion ethically questionable".

As has been mentioned throughout this dissertation, many ethical concerns with regard to sexbots can, at best, be speculative given that sophisticated sexbot technology is in its infancy. However, as far as social isolation is concerned, it must be noted that there has been concern in Japan about falling birth rates that may be due to men's preference for virtual girlfriends as opposed to human girlfriends. A 2010 survey conducted by the Japanese Ministry of Health, Labor and Welfare noted that this preference might be due to human partners being much more complicated and demanding than virtual partners (Rani 2013). One could extrapolate from this and realistically consider if this may potentially be the case with sexbots, especially given that interaction with them may be even more realistic, but without the complicated and demanding nature of human relationships.

This would only be an issue, however, should meaningful interaction with a sexbot be 'less than' compared to meaningful interaction with another human being. We already established that meaningful interaction, and particularly meaningful interaction in the form of a romantic lovelike relationship with a sexbot, would differ from such a relationship with another human being. The question, however, is *why* love as experienced with another person is better than love as experienced with a sexbot. Nyholm & Frank (2019) hold that a romantic relationship with another human being would be much richer, fuller, more multi-dimensional, and more valuable than a relationship with a sexbot in which the human interactant treats their sexbot as if it were a human partner. We could consider this being the case given that, as pointed out earlier, love as experienced with a human being can at least be a mutual love, as opposed to love as experienced with a sexbot which is functional in nature and, as such, non-reciprocal and, therefore, not a mutual love. As Danaher (2018) states:

> "Philosophers emphasise the need for mutual commitment in any meaningful relationship. It's not enough for you to feel a strong, emotional attachment to another; they have to feel a similar attachment to you. Robots might be able to perform love, saying and doing all the right things, but performance is insufficient".

Thus, the quality of the kind of love as experienced with a sexbot is questionable, and we can therefore be sceptical about the notion of a sexbot being an adequate substitute for a human partner (see e.g. Nyholm & Frank 2019). We could consider love as experienced with a sexbot as impoverished as compared to love that can be experienced with another human being.

An added concern is that should we become accustomed to meaningful interactions and romantic relationships with sexbots, thus disinclining us to interact and bond with other people, this could fundamentally change our impression of what it means to bond with another human being in a romantic way. Becoming accustomed to a somewhat impoverished relationship with a sexbot (as compared to another person) may impact human interactants in such a way that we come to associate romantic relationships as ones that are superficial and one-dimensional, thus, human interactants may develop the impression that "this is all there is and could be to a relationship" (Nyholm & Frank 2019). "In this way, the robot relationship might not only block off the instantiation of human relationships, but also the very conception thereof" (ibid.).

Of course, there may be instances wherein some people are in a situation where they are, or maybe wish to remain, socially isolated from other people. Think of the case of a person who is physically disabled and who may struggle to go out and interact with other people. Or someone who is too painfully self-conscious to interact with other people, and thus may prefer to not socially interact with others. In such instances, we could contend that some type of social interaction – even limited social interaction with a sexbot – would be better than no social interaction at all. This may be the case, however, my focus, as mentioned at the beginning of this section, is on the problematics that may arise should social interaction with a sexbot (and functional love) actually come to replace social interaction (and mutual love) with another human being. Therefore, I do not think that social interaction with a sexbot for someone who is already extremely socially isolated would necessarily be as morally and socially problematic as someone who is able to socially interact with and experience mutual love with another human being, but chooses not to. As Sullins (2012) most aptly states:

> "What I am saying instead is that we have to be careful to not mistake simulacral love for the real thing. I agree that this technology will be compelling, and in fact already is compelling to some early adopters. But the kind of relationships that are evolving are not philosophically erotic, that is challenging and compassionate, but rather one-sided affairs over burdened by fleeting passions and the desire to erase everything in the beloved that is not a complete reflection of the lovers preconceived notions of what he or she thinks they want out of a partner".

The different nature of this love would also fundamentally change us as moral beings. Our meaningful interaction with our sexbot, particularly if we are socially isolated, may prevent us from practicing certain moral skills. This takes us to the next section wherein I discuss how meaningful interaction with sexbots may lead to "moral deskilling" (see e.g. Vallor 2015).

3.2.2. "Moral deskilling"

The argument that the integration of new technology into society could contribute to "moral deskilling" in society is one put forward by Vallor (2015) in the article *Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character*. Her insights and argument are particularly important to consider in this context given that many issues concerning human morality in relation to social robots relate to how our immoral treatment of them may negatively impact our morality, thus negatively impacting the way in which we treat other people (as was discussed in Chapter 3 and 4). However, drawing upon Vallor's (2015) work, we can consider the way in which treating sexbots *morally well*, and meaningfully interacting with them, may also negatively impact our morality, and our moral treatment of other people.

The concept of "deskilling" has been "used to frame the way in which twentieth century advances in machine automation resulted in the economic devaluation of practical knowledge and skillsets historically cultivated by machinists, artisans, and other highly trained workers" (ibid.: 108). Vallor, however, uses the concept of deskilling in relation to morality, and particularly moral skills.

> "… moral skills appear just as vulnerable to disruption or devaluation by technology-driven shifts in human practices as are professional or artisanal skills such as machining, shoemaking, or gardening. This is because moral skills are typically acquired in specific practices which, under the right conditions and with sufficient opportunity for repetition, foster the cultivation of practical wisdom and moral habituation that jointly constitute genuine virtue. The driving concern of this paper is that profound technological shifts in human practices, if they disrupt or reduce the availability of these opportunities, can interrupt the path by which these moral skills are developed, habituated, and expressed" (ibid.: 109).

122

Vallor (2015) discusses three instances in which moral deskilling could be hazardous: the use of autonomous weapons systems, new media practices and multitasking, and robot caregivers.

As far as autonomous weapons systems are concerned, many suggest that human soldiers are already, or may soon be, "the weakest link in the military arsenal" (Heyns 2013: 10). Given this possibility, Vallor (2015) is concerned that human soldiers may become morally deskilled. This is because, as Vallor (2015: 114) states: "The right use of military force meets our test of a moral skill; force is something that must be applied only at the right times, towards the right persons, in the right places, and in the right manner". It is only through active participation in war that soldiers can practice such a moral skill. As such, although human beings would still be responsible for the starting, funding, and continuation of wars, we may not know how to morally execute them.

Concerning new media practices and multitasking, Vallor (2015) is concerned that we are losing the skill of paying attention. Chronic multitasking may negatively impact cognitive abilities, thus making us more easily distracted, and less able to refocus our attention once it has been lost (ibid.: 117). Paying attention is an important moral skill to have:

> "A person who cannot be counted on to pay attention when you tell her about the recent death of your closest friend, or who is unable to stay focused on the grave and imminent danger to which you're trying to alert her, or who cannot attend to the expressions on your face during an intense conversation, is not someone who can be said to be virtuous" (ibid.).

Lastly, with regard to robot caregivers, Vallor (2015) is concerned that robotic caregivers will lead to us losing the skill of caring, or at least caring effectively (this was discussed in detail in Chapter 2). Caring is an important moral skill because "[i]t is difficult to know how to care for people well—emotionally, physically, financially, and otherwise, in the right ways, at the right times, and for the right persons" (ibid.: 119). Should robots replace caregivers, Vallor (2015) is concerned that caregivers will not practice the skill of giving care, thus becoming morally deskilled in this regard.

Drawing upon Vallor's (2015) ideas about moral deskilling in relation to the abovementioned technologies, we could also consider the issue of moral deskilling in relation to meaningful interaction with sexbots.

Let us consider how sexbots will literally be made for us; designed to our ideal physical preferences, and to fulfil our every need. They may be able to fulfil our every need not only because they may be preprogrammed to do so, but because machine learning will allow them to continuously learn about our needs, and the adequate ways in which to fulfil them. Moreover, as discussed in the previous section, they will be fully committed to doing so. Sullins (2012: 16) questions whether people may "grow less and less tolerant of human relations", as we may lose our "tolerance to deal with others who are not preprogrammed to serve our every need". The first moral skill that comes to mind in this context is that of patience. As the well-known phrase states "patience is a virtue". It is a skill that is practised in everyday interaction with other human beings and, as such, also practised when interacting when intimate partners. Perhaps we would become so accustomed to our sexbot being there at our every beck and call, fully committed to fulfilling our needs in a lovingly committed way, that we will forget how to exercise the skill of patience. In losing this skill, we may become less patient in our relations with other people who, for various reasons, cannot be there at our every beck and call the way our robot can be, thus causing us to be less tolerant of other people. Human partners, although they may share their lives with us, and care about our wellbeing are people with their own needs and desires. As Sullins (2012) states:

> "Fellow humans … represent a much more difficult problem and do not always readily change to accommodate one's every need. They provide resistance and have their own interests and desires that make demands on the other person in the relationship".

This would not be the case with a sexbot which does not have its own needs given that its only 'desire' is to serve you – this is what they are designed to do and to be: your ideal partner.

This problem would link back to issues surrounding social isolation: should we become less tolerant of other people, we may become disinclined to meaningfully interact with them. As Sharkey et al. (2017: 21) state: "…relationships with robots are fictive and may decrease our ability to interact with other humans". This would be a vicious cycle wherein because we are disinclined to interact with other people due to our losing the skill of patience, for instance, we would become socially isolated. This social isolation would therefore not provide ample opportunity to practise the skill of patience, thus only compounding the situation.

More than moral deskilling, drawing upon Wollstonecroft (2009), we could also consider whether relationships with sexbots could actually cultivate vices in human interactants (Nyholm & Frank 2019). According to Wollstonecroft (2009), unequal relationships breed vices within the person that has the higher status. Examples of such vices include self-centeredness and a tyrannical and domineering personality. As was noted above, meaningful interaction between a human interactant and a sexbot is inherently asymmetrical or unequal and, therefore, there may be the possibility that meaningful interaction with a sexbot could not only morally deskill us, but actually cause human interactants to develop certain vices.

Given the moral issues surrounding the actual creation of sexbots that are designed to manipulate our psychological tendency to anthropomorphise, as well as moral issues that may arise in relation to human interactants who may choose to 'meaningfully' interact with, and love, their sexbots as opposed to being in a romantic/loving relationship with another human being, I will now suggest ethical conditions for human-sexbot interaction that we may consider in relation to the roboticists who design sexbots, as well as the human interactants. I hold that both parties are morally responsible for the repercussions of 'meaningful' interaction with sexbots, but in different ways.

4.   Ethical conditions of 'meaningful' human-sexbot interaction

Given the possibility that meaningful interaction with sexbots could have negative moral impacts upon human interactants and society at large, it is important to address ways in which we could limit these impacts. This is especially the case if we are particularly concerned with the possibility of sexbots actually substituting human partners. I discuss two ways in which we could address this issue: from the perspective of the roboticists who design and create sexbots, and from the perspective of human interactants. Both perspectives raise important considerations, however, I hold that it is ultimately the responsibility of human interactants to ensure that sexbot technology does not have deep, widespread negative impacts upon humans themselves and the society in which they live. As such, it remains the case that we must maintain an anthropocentric perspective when investigating ethical conditions of human-sexbot interaction.

4.1. Responsibility of roboticists

We may have to consider the way in which sexbots are designed and marketed. If there is concern that the human-like appearance of sexbots could too easily deceive human interactants into treating them as if they are human, or even subconsciously believing that they are human (Scheutz 2012: 215), then perhaps they should not be designed to look so realistically human, and behave in such realistically human ways?

Sullins (2012) notes the difference between a "variance" design strategy and a "mimesis" design strategy. The variance design strategy is one in which "machines would be able to simulate emotion and embodiment in the machine in a way that is not a direct imitation of the human though it may be inspired by the study of human emotion" (ibid.). As such, there is a "certain distance maintained between the robot and the expectations of the user, they look and act more or less like machines so the user expects less human verisimilitude from them". The mimesis design strategy, however, is one in which the goal is to make a "much stronger appeal to the user's emotions in order to have the user treat the machine as if it were a fellow human agent or at the very least to have the user be momentarily confused as to whether or not the robot is a human or an android and is the strategy we can argue is employed in the Roxxxy sex robot as well as the much more complex Geminoid[48] androids" (ibid.). Those who support the latter design strategy hold that if the intention is to create robots with which we can come into close, intimate contact, then these robots must necessarily be almost indistinguishable from human beings.

We could, perhaps, vouch for the former design strategy by way of arguing that this lessens the likelihood for human interactants to anthropomorphise to such an extent that they would see nothing wrong with substituting a human partner with such a sexbot. We could also, possibly, consider, as Boden et al. (2017) suggest, that there should be a way for human interactants to "find out what [their sexbot] really is" so that human interactants can be made aware that their sexbot's intelligence is artificial. Nyholm & Frank (2019) relate this to the way in which Sophia is designed, for example:

> "Think, for example, of the back of the head of the famous robot Sophia from
> Hanson robotics. While Sophia's face looks very much like a human face, the

---

[48] "A Geminoid is a special type of android that doesn't just broadly resemble a human; it is meant to look like a specific person" (Robohub 2015).

back of Sophia's head is literally transparent and highly robotic in its appearance, so that it becomes obvious that Sophia is a robot".

Such a suggestion is, however, problematic for two reasons: firstly, there is the possibility that designing sexbots in such a way may mean human interactants could be less likely to engage with them in an emotional way – and is this not the point of a sexbot? Is the point not for human interactants to interact with them the way in which one may interact with a human partner (albeit in a unidirectional way)? As Nyholm & Frank (2019) state: "[the] kinds of cues and other means that would help to engage users' emotions – e.g., facial expressions and things the robots would 'say' – would presumably work best if they are designed to be human-like rather than maximally robot-like" and "[to] the extent that the designers make the machine nature of the robot explicit or obvious, this presumably diminishes the appeal of the robot as an entity someone might want to spend time with, have sex with, and even express affection towards".

Secondly, I am doubtful that this would necessarily make a difference. As was mentioned above, if a human interactant is desperate enough to believe something, they will allow themselves to be deceived. Moreover, specifically as far as anthropomorphisation is concerned, people anthropomorphise at varying degrees (Waytz, et al. 2014: 226) and, moreover, for different reasons. This makes it difficult to see how people's perspectives of robots could be hugely impacted via design strategies. There is anecdotal evidence of people already anthropomorphising or even having close relationships with current robots that are nowhere near being indistinguishable human simulacra, such as with their AIBOs, or even Cynthia Breazeal and her Kismet robot[49].

The tendency to anthropomorphise is the result of one's trying to "rationalise an entity's behaviour in a given social environment" (Duffy 2003: 180) as pointed out in Chapter 1. It is hypothesised that the tendency to anthropomorphise evolved, because it

---

[49] "Cynthia Breazeal was leader on the design team for Kismet, the robotic head that was designed to interact with humans 'sociably', much as a two-year-old child would. Breazeal was its chief programmer, tutor, and companion. Kismet needed Breazeal to become as 'intelligent' as it did and then Kismet became a creature Breazeal and others could interact with. Breazeal experienced what might be called a maternal connection to Kismet; she certainly describes a sense of connection with it as more than 'mere' machine. When she graduated from MIT and left the AI Laboratory where she had done her doctoral research, the tradition of academic property rights demanded that Kismet be left behind in the laboratory that had paid for its development. What she left behind was the robot 'head' and its attendant software. Breazeal described a sharp sense of loss. Building a new Kismet would not be the same" (Turkle 2006).

"favoured cooperation among early humans" and improved human fitness by helping us "distinguish between friends and enemies, to rapidly recognize predators, and to establish alliances with other tribes" (Damiano & Dumouchel 2018). As such, it is an evolutionary trait inherent within us all. This being the case, I find it hard to believe that simple design strategies can possibly prevent anthropomorphisation from occurring when it is so deeply engrained into the human psyche, i.e., something that "constitutes a fundamental and permanent dimension of the human mind" (ibid.).

It has also been suggested that designers could be transparent by way of not only making the appearance of social robots more transparent, but by way of "warning labels" such as the warning labels on cigarette boxes (see e.g. Nyholm & Frank 2019 and Scheutz 2012). This could be done by way of the sexbot playing messages to human interactants every now and again that remind them that they are robots, and are in no way actually human beings and, therefore, function differently. Nyholm & Frank (2019) state:

> "The sex robot could say things such as 'remember that I am a robot and that humans and robots differ in how they function' or 'recall that as a robot, I operate in a very different way than a human does'. And the robot could be made to say such things every now and then, not just the first time the user interacts with the robot. The socially interactive sex robot could be made to say such things often enough that the user would not forget it. In addition, the user could sometimes be asked to reaffirm their consent to using a robot that is programmed to elicit social responses. The robot could say, for example, 'now that I have reminded you that I am a robot and that humans and robots function in very different ways, do you still consent to the interaction that you have with me?' or messages like that".

However, I think that we would still run into the same issue as discussed above. Human interactants could disregard this and still allow themselves to be deceived should they so wish to be.

Although I am sceptical that design specifications would solve the problem of people bonding to their sexbots to such an extent that they become replacements for human partners, this does not discount the value of this consideration. Rather, we must look at it not from the perspective of roboticists having the sole responsibility to prevent potential moral and social harms from happening (such as those discussed in the section before this), but having a

128

responsibility in tandem with the responsibility of human interactants. Thus, roboticists must be transparent about the true nature of sexbots but not so as to solve potential moral and social harms, but so that human interactants can make informed decisions about how they want to interact with their sexbots, and how much they integrate them into their lives. As such, the next section discusses how and why human interactants should also be responsible for potential moral and social harms that 'meaningful' interaction with sexbots may bring about.

4.2. Responsibility of human interactants

If there is concern that transparent design would not necessarily prevent moral and social impacts given the strong human tendency to anthropomorphise, we may question whether the very existence of this technology can be a good thing. However, we must not forget the very real positive impacts that this technology can have (as discussed in Chapter 2). And we must also go back to my discussion in Chapter 2 about why it is important to consider an anthropocentric lense in relation to analysing positive and negative impacts of technology. Verbeek (2011) calls humans "technomoral creatures": technology has, and continues to, impact us morally. There is the possibility for technology to impact us morally in both positive and negative ways. However, it remains the case in this context of meaningful interaction that human interactants remain responsible for the moral changes this technology may bring about. As such, the responsibility is not for roboticists to *prevent* negative impacts, or ensure positive impacts; their responsibility is to be transparent as far as the nature of sexbots is concerned, so as to allow human interactants to be informed about the nature of the technology with which they interact, thus allowing them to make informed decisions about how to integrate this technology into their lives, and whether they allow this technology to morally impact them and, if so, in positive or negative ways. Being informed would at least allow human interactants to make responsible decisions as far as their use of, and interaction with, sexbots is concerned.

Drawing upon the work of Verbeek (2011), Vallor (2015: 116) states:

"We still think of technologies as neutral tools for accomplishing things, rather than as systems that merge with, mediate and continually transform our own agency and capacities ... Thus our first thought is always 'How can I use this new, more powerful technology to do something for me?' Very rarely is

129

our first question, 'How should this new technology be related with me?' or more precisely, 'How can I relate to this technology in a way that helps me become the sort of human being I would like to be?'".

Vallor (2011) thus draws attention away from the way in which technology itself impacts our lives, and towards considering the way in which our relation with technology impacts us as moral and social beings.

As such, both roboticists and human interactants are responsible for the moral and social impacts of sexbots. However, human interactants make the ultimate decision and they have the liberty to do so. As Mill (1986) stated: "The only part of the conduct on anyone, for which he is amenable to society, is that which concerns others. In the part which merely concerns himself, his independence is, of right, absolute. Over himself, over his own body and mind, the individual is sovereign".

As such, and in accordance with Whitby (2012: 239), if human interactants want to have meaningful interactions with their sexbot in such a way that they allow themselves to become socially isolated and morally deskilled, then they absolutely have the right to do so. However, such a decision would at least have to be an informed one: they know these are the risks, yet they have chosen to live with it. Much in the same way that smokers are informed on their box of cigarettes that smoking causes lung disease, yet they make an informed decision to continue smoking. As Vallor (2015: 118) states:

> "The choice is not between surrendering to technology or liberating ourselves
> from it. We are technomoral creatures to the core; that is, we allow and have
> always allowed the things we make to reshape us. The only question is
> whether this process is deliberate and wise or unreflective and reckless".

Of course, should every human interactant decide to substitute their human relationships, with a 'relationship' with a sexbot, such a change would not only impact the life of that individual human interactant, but the kind of society in which we live and have come to understand would fundamentally change. Moreover, there would also be widespread societal change if human interactants all gradually became morally deskilled in the sense discussed in section 3.2.2, and we gradually came to start treating fellow human beings in less moral ways. In this case, we may have to consider whether widespread 'meaningful' interaction with sexbots would cause harm to society at large. Thus, although human interactants would have the liberty to individually decide if they want to allow sexbot

technology to negatively impact them morally and socially, it may read further than this. As such, we would have to weigh up whether concern surrounding these societal impacts is more important than individual liberty. As Whitby (2012: 240) states:

> "The immediate conclusion urged is that the availability of robot lovers and caring technologies raises these political debates and should be discussed in the political area, rather than simply as technology. There is a need for more scientific research into these social effects. There is also a need for balanced general public debate on the moral question of whether or not such social effects are to be held more important than individual liberty".

However, it would still remain the case that human interactants hold responsibility in this regard in the sense that we must educate ourselves about the nature of this technology, as well as the impacts it may have, and be consciously aware of the ramifications of our decisions to interact with this technology, thus being able to have informed conversations about whether societal impacts outweigh individual liberty or not.

5. Conclusion

This chapter addressed the '*how'* distinction of our interaction with sexbots from a 'positive' perspective i.e. the implications of treating sexbots *morally well*. Given the subjectivity of what would entail the 'moral' treatment of a sexbot, I specifically focused discussion on 'meaningful' interaction with sexbots in the context of the possibility of experiencing romantic 'love' with a sexbot. I argued that human interactants can experience love with a sexbot, but that this love would be different in nature compared to love as experienced with another person: it would be functional in nature, as opposed to mutual. However, regardless of the (im)possibility of love with a sexbot, it remains the case that human interactants can *perceive* their sexbots as reciprocating love, thus it is important to consider the implications that may arise should human interactants choose to replace human partners with sexbot partners.

I considered the ethics of 'meaningful' human-sexbot interaction from two perspectives: that of the roboticists and that of human interactants. As far as roboticists are concerned, I discussed issues surrounding the concept of deception: that the very creation of sexbots designed to deceive human interactants may be considered ethically dubious.

However, I also argued that human interactants play a role in this deception as they allow themselves to be deceived. As far as human interactants are concerned, I discussed issues of social isolation and 'moral deskilling'.

Given these ethical concerns in relation to roboticists and human interactants, I put forward suggestions as to how these implications could be mitigated from the perspective of roboticists and the consideration of design specifications, as well as human interactants and the role they play in being responsible for making informed decisions as to whether they allow their interaction with sexbots to negatively impact them or not. Thus, ultimately, the onus is, once again, on human interactants to ensure that sexbot technology is beneficial to them, and to society at large.

Conclusion

This dissertation sought to conduct a philosophical investigation into the ethical conditions for human-sexbot interaction. It did so by way of investigating the moral problematics that may arise in relation to our interacting with sexbots in terms of *what* we may use them for, as well as *how* we may use them, and then put forward possible conditions that we may consider as a way to try and ensure that sexbot technology is mainly beneficial to us as human interactants. It emphasised the importance of maintaining an *anthropocentric perspective* with regard to both understanding the way in which sexbot technology may impact our lives, as well as the possible conditions that we may then consider so as to prevent moral and social harms.

As such, the main take away from this dissertation is that when considering whether sexbot technology is beneficial to society or not, thus possibly impacting decisions about whether it is a good idea for this kind of technology to be designed, created, and made available for use, we cannot only focus our attention on the technology itself, thereby deeming the technology as essentially 'good' or 'bad' for society. It is not the technology itself, which will deem whether its integration into society is beneficial or detrimental. Rather, their being beneficial or detrimental to society is dependent upon *our relation* to sexbots: it is how we choose to interact with, or treat sexbots, that determines whether they will positively or negatively impact our lives morally and socially. Thus, we are the masters of our moral fates. This is the case with technology in general, as was discussed in Chapter 2 and 5: "[T]echnology is never of itself good, or bad, or neutral. It is always all three, depending on how we use it" Whitby (2012: 237).

There are very real ways in which sexbots can benefit society, but in order to ensure that this technology remains beneficial, the onus is on *us* (human interactants) to be aware that the way in which we interact with them can affect whether they are beneficial to us or not. If we are not aware of this, the negative impacts may be great, thus calling into question whether their very existence is a good idea. However, were we to question their very existence due to the fear of negative implications of their use, without taking into consideration the role that human interactants play in determining whether these implications are good or bad, I believe we would only be detrimenting ourselves because although we would be preventing negative impacts of sexbots in our lives, we also would then not have the opportunity to experience very real positive impacts of our interaction with them.

I drew this conclusion by way of investigating: the nature of sexbots and our interaction with them, the impacts upon human interactants depending on *what* we use them for as well as *how* we may use them, and possible conditions we may consider for human-sexbot interaction – specifically in the context of *how* we use sexbots – so as to try and ensure that this technology is beneficial to us. This was done in five chapters that dealt with different aspects of these main points.

Briefly, Chapter 1 lay down the conceptual groundwork needed so as to understand the kind of technology about which I speak, thus shedding light on why it is possible for human interactants to interact with sexbots in such a human-like way that makes it important for us to consider the very real moral and social ramifications that may arise in relation to our interaction with them.

Chapter 2 then made a very important conceptual distinction between *what* we use sexbots for and *how* we use sexbots. This distinction separated the dissertation into two parts wherein Chapter 2 specifically looked at the positive and negative impacts of *what* we may use sexbots for, and Chapters 3-5 then investigated the implications of *how* we may use sexbots and possible ethical conditions we may consider in relation to how we use them. Given this conceptual distinction, Chapter 2 introduced us to my reasoning for why we must take on an anthropocentric perspective when we consider the ways in which sexbots may impact us, as well as ethical conditions that we may consider in relation to our interaction with them.

Chapter 2 then went on to exemplify why this anthropocentric perspective is important because it means focusing more on *how* we use sexbots as opposed to *what* we use them for. By showing how, in each aspect of *what* we may use sexbots for, there are both potential positive and negative impacts, it sheds light on the point that it is not what sexbots essentially are, or what they can be used for, that makes them good or bad. Rather, we should focus on *how we interact with them* i.e. *how* we use/relate to them, regardless of *what* we may use them for.

This is an important consideration because should we only focus on what they may be used for, home in on the negative aspects, deem them too great, and then call for the development of sexbots to be halted or limited, then human interactants would not only be protected from negative implications, but would be prevented from experiencing their positive impacts too. Given this, the rest of the dissertation shifts away from focusing on the capabilities of sexbots (i.e. the *'what'* distinction), and toward the human interactants who

134

interact with them, particularly looking at how we interact with them in both a positive and negative sense (the *'how'* distinction).

The rest of the dissertation focuses on a way in which sexbot technology could remain beneficial to society. The topic of beneficial AI, and how to ensure it, is a vast one. In terms of considering how to ensure sexbots in particular are beneficial to society, this is also a big task, especially given that the technology is in its infancy, and concerns about the integration of sexbots into society remain speculative. However, I put forward that at least one way in which we could go about tackling this difficult task is focusing upon our relation to the technology, i.e. investigating *how* we may interact with sexbots, the moral ramifications thereof, and how we may consider preventing negative ramifications by way of ethical conditions for human-sexbot interaction considered from an anthropocentric perspective. This is what Chapters 3 to 5 focused on.

Chapter 3 investigated the moral implications that may arise should human interactants treat sexbots in *immoral* ways, thus looking at *how* we may use sexbots in a 'negative' sense. It is argued that treating sexbots immorally would damage the moral fibre of human interactants and, therefore, human interactants should not treat sexbots immorally. I put forward that we should consider certain ethical conditions for beneficial human-robot interaction, in relation to our immoral treatment of sexbots, not because sexbots warrant moral treatment, but because human interactants need protection from their own potential immoral behavior towards sexbots. Thus, we should consider ethical conditions for beneficial human-robot interaction from an anthropocentric perspective.

However, calling for human interactants to not treat sexbots immorally is merely a philosophical suggestion, and, therefore, the issue arises whether there may be a way in which we could ensure that such a suggestion could have real-world applicability. Given this concern, in Chapter 4, I suggested that one way, in which we could ensure this, would be to consider granting sexbots negative rights. Although the focus here may initially seem 'robocentric' as opposed to anthropocentric, my reasoning for granting sexbots negative rights is for *our own good*, not theirs, thus the anthropocentric focus remains – *our* behavior towards sexbots should be controlled for *our* moral benefit.

After focusing on human-sexbot interaction in a negative light in Chapters 3 and 4, Chapter 5 then considers the moral impact of *how* we may use sexbots in a 'positive' sense in terms of our 'moral' treatment of sexbots in the context of considering 'meaningful' interaction with sexbots, particularly in terms of the potential to experience romantic 'love' with a sexbot. Given moral and social problems that may arise in this regard, I argue that both

135

roboticists and human interactants have a responsible role in ensuring that 'meaningful' interaction with sexbots does not cause moral and social harms. However, I argue that it is still human interactants who are ultimately responsible in terms of ensuring that sexbot technology does not morally or socially harm them and society at large.

Given this overview of the arguments I have made in this dissertation, there are two potential critiques that I will raise, and which I will address: firstly, that my dissertation seems to have a particularly negative focus on the implications of sexbots in terms of *how* we interact with them. Secondly, that the consideration of ethical conditions in relation to human-sexbot interaction is morally paternalistic.

As far as the first critique is concerned, I do acknowledge that Chapters 3 and 5, in terms of considering immoral and moral treatment of sexbots, only focuses on negative moral implications that may arise. My reasoning for this focus, however, is not to cast human-sexbot interaction in a bad light – I have acknowledged the ways in which sexbots may benefit society in many places, specifically in Chapter 2.

Rather, in an indirect way, my focusing on the possible negative implications of human-sexbot interaction does contribute to our understanding of, and possibly ensuring, that sexbot technology is beneficial. Although there are real positive impacts, this does not require as much ethical investigation given that I find it hard to believe that anyone would take issue with the integration of technology that fundamentally makes our lives better. In an indirect way, being aware of the problems that this technology may bring about means ensuring that the technology is good for us. If we are aware of the potential negative ramifications, then we can focus on preventing these from occurring and, as such, hopefully, the positive ramifications will remain while the negative ramifications are avoided.

Secondly, my suggestion that we should consider ethical boundaries in relation to our interaction with sexbots may be considered morally paternalistic. After all, as was mentioned in Chapter 5, should human interactants not have the individual liberty to decide how they interact with sexbots, and whether they allow their interaction with them to cause them individual moral harm? This would potentially be a critique especially as far as Chapter 4 is concerned, wherein I suggested sexbots be granted negative rights so as to prevent human interactants from treating them immorally.

"Paternalism is the interference of a state or an individual with another person, against their will, and defended or motivated by a claim that the person interfered with will be better off or protected from harm" (Dworkin 2020). Paternalism particularly arises in relation to restrictions made in law such as not being allowed to drive without a seatbelt, various anti-

136

drug legislations, or legislations that criminalise prostitution. When the state or an individual interferes with another person for the moral wellbeing of that person, then this is a case of moral paternalism. Moral paternalism may arise for two reasons: "interferences to improve a person's moral character, and hence [their] well-being, and interferences to make someone a better person—even if [their] life does not go better for [them] as a result" (Dworkin 2020).

Given the suggestion that we should grant negative rights to sexbots so as to prevent human interactants from treating them immorally thereby preventing the moral fibre of humans from being negatively impacted, one could contend that this is morally paternalistic towards the interactants themselves. It is a suggestion that, through legal means, we interfere with the way in which people may want to interact with sexbots through preventing them from treating them in a certain way. This is so as to improve (or at least not negatively impact) that person's moral character. It is a hindrance upon the freedom of the human interactant to treat their robot in any which way they please and, particularly, the freedom of being able to treat them immorally should they wish to do so.

Many arguments have been made that paternalism is wrong when it infringes upon a person's autonomy, which it commonly does. Being made to drive with a seatbelt, for example, infringes upon my choice to drive without one. As stated by Birks (2018):

> "It is a commonly held liberal view that if a person is making a self-regarding autonomous choice, it is always all things considered wrong to interfere paternalistically with this choice, irrespective of the benefits of the interference to the person's life. This view is often attributed to John Stuart Mill, and has been defended by many philosophers, such as Richard Arneson, Michael Cholbi, John Hodson, Joel Feinberg, Jeffrie Murphy, and Donald VanDeVeer".

Aside from general arguments that moral paternalism is wrong, because is restricts people's freedom, another possible objection is one discussed by Scoccia (2008: 377). The claim here is that prohibiting immoral acts does not weaken the bad character traits that those immoral acts express, because although we can prohibit people from acting in a certain way, thus influencing their actions, laws cannot ultimately alter the motivations that are behind those acts. For example, Scoccia (ibid.) states: "A ban on prostitution can deter a man from frequenting prostitutes by exploiting his fear of discovery and arrest, but it will not thereby make him either less intemperate or less disrespectful of humanity".

137

I, however, argue that the suggestion of considering ethical conditions in relation to human-sexbot interaction by way of granting sexbots negative rights, is not problematic with regard to concerns surrounding moral paternalism. As far as the rejection of moral paternalism on the ground of it restricting people's freedom is concerned, freedom may be restricted in the context I am focused on, but it would only be minimally so. This is because I am not here prescribing in detail how people should treat their robots but am rather just suggesting that they at least should not treat sexbots immorally. As such, there may be *some* interference with autonomy in the sense that one is restricted from treating sexbots immorally should they wish to, but I am not restricting them from making other choices in relation to their sexbot such as whether to take their sexbot companion out on a date, or ask it how its day was when they return home from work.

One can compare this to laws in relation to freedom of speech or public indecency, for example. I have freedom of speech but am restricted from hate speech, and I can wear whatever clothes I choose, but laws against public indecency prevent me from walking around in the nude. In both instances, there is some restriction to my freedom but I am not entirely restricted. And if some restriction in instances such as these is beneficial to both the moral agents themselves and general society in the long run, as I suggest here may very well be the case, then I do not see placing some restrictions in the form of ethical conditions concerning our treatment of sexbots, such as the consideration of negative rights for sexbots, as problematic.

As far as Chapter 5 is concerned, the issue of moral paternalism would not arise given that I suggest ethical conditions in relation to human interactants taking responsibility for their own actions, in terms of making informed decisions as to whether they allow meaningful interaction with a sexbot to cause them moral harm. As such, I did not argue that there should be explicit interference in the freedom that human interactants have to treat sexbot morally well such as by meaningfully interacting with them, i.e. experiencing 'love' with them. As was discussed, the concept of what the moral treatment of sexbots would entail is subjective and, as such, it was argued that human interactants should have the freedom to 'meaningfully' interact with their sexbot – such as 'love' them – whatever this may entail for them. However, human interactants should at least be made aware of the nature of sexbots (by roboticists) so that their decision to interact 'meaningfully' with a sexbot is a fully informed one.

We may also question why I explicitly suggest more stringent ethical conditions for the immoral treatment of sexbots, but not for moral treatment, despite there being negative

implications in both cases. My reasoning for stringent ethical conditions for immoral treatment lies in my concern that the very act of treating a sexbot immorally is ethically dubious (this was argued in Chapter 3). It is symbolic of treating a human partner immorally. As such, I do strongly feel that there should be ethical boundaries in this regard. As far as moral treatment is concerned, there is nothing inherently wrong with treating a sexbot morally well, such as in the context of experiencing 'love' with it, as was argued in Chapter 5. Therefore, I do not think moral treatment in this regard should necessarily be prevented or restricted. It should rather be a case of people having the freedom to 'love' their sexbots, with the only condition being that they should be aware of the potential implications thereof for themselves (and perhaps society at large). Ethical conditions for this kind of interaction should be considered in the form of transparency from roboticists, thus allowing human interactants to ultimately make informed decisions as to whether they allow meaningful interaction with sexbots to cause them moral harm or not.

Having addressed potential critiques to my arguments, let us consider some suggestions for future work.

Concern was raised in Chapters 3 and 4 that we do not have a clear understanding of what treating a robot 'immorally' or 'morally' would essentially entail. I worked around this issue by way of defining immoral behavior in terms of what we would consider as immoral treatment of another human being. I argued that we could do this given that sexbots are designed and created to be human simulacra, thus the way we interact with them is symbolic of interaction with another human being. As far as 'moral' treatment is concerned, I argued that moral treatment is far more subjective in terms of human-human interaction, particularly in the context of romantic relationships; hence my focus on the concept of 'functional love' with sexbots.

However, I feel that continuing to work around these conceptual issues would not allow for questions concerning conditions of human-sexbot interaction to move forward, particularly in terms of real-world applicability in the context of the law. Work needs to be done that will help to clearly define what would constitute moral/immoral behavior specifically in the context of human-sexbot interaction. Having an understanding of this would not only help human interactants have a clear understanding of conditions they should personally consider when interacting with a sexbot, but will also help should we really consider the granting of rights to robots as Chapter 4 suggests. We could only do so should we have a clear understanding of immoral treatment that could be legally defined, thus allowing human interactants to suffer legal implications of their misdeeds towards sexbots.

Having raised the issue of real-world applicability of certain conditions by considering robot rights, we may then consider whether this is truly implementable. As is the case with rights in general and, specifically in this case with regard to negative rights, the onus is on the moral agent to see that they abide by the law. Of course, there are policing systems in place to oversee that laws are generally abided by, but policing systems cannot be everywhere at once to ensure that this is happening. Hence, there is a responsibility for the moral agent to abide by the law, and there is a responsibility for the moral patient to report any actions towards them that have infringed upon their rights.

In terms of the responsibility of the moral patient, consider the following: a victim of physical abuse (the moral patient) for example, would have the responsibility to report their abuser (the moral agent) to the police so that the abuser would be held accountable for breaking the law. Could there be a similar moral patient responsibility for robots? Could there be a system in place whereby the mistreatment of sexbots could be reported to relevant authorities by the sexbots themselves? There would have to be some kind of system in place that would deter human interactants from treating robots immorally otherwise, practically speaking, I do not think this suggestion could work if human interactants could not actually be held accountable for their moral actions.

Holding human interactants accountable would require contributions made by not only AI ethicists, but computer technologists as well as legal theorists. Especially as far as AI ethics and the law is concerned, there is definitely space for more contribution from legal philosophy in this regard, and also perhaps with disciplines such as anthropology and of course psychology.

Despite more research being needed in these regards, this dissertation has contributed to the field of AI ethics and HRI in general, and specifically to human-sexbot interaction, in two ways. There is a lot of work that investigates the positive and negative aspects of sexbots, specifically calling our attention to how important it is to be aware of the negative impacts. However, research is lacking in terms of actually putting forward solutions as to how we could prevent these negative impacts. Where there are suggestions, it focuses more on the technology itself – whether or not it should be created, whether it should be allowed to be implemented, how it should be designed – yet there is little that is concerned with the role that human interactants play in the materialisation of negative ramifications. Thus, my anthropocentric focus adds an important perspective that is sorely lacking in ethics of sexbot literature.

Moreover, in the suggestion of granting negative rights to robots, a contribution is not only made in terms of sexbot literature, but social robotics and robot rights in general. Literature is divided on whether rights should be granted to robots as if it is an all or nothing situation, without considering the kind of rights that could be considered. As was mentioned, although the consideration of negative rights is still a broad categorisation, it is a categorisation that I have not as yet come across.

My emphasis on an anthropocentric perspective is not only important in the context of human-sexbot interaction, but our interaction with technology in general. Although interaction with sexbots may have moral impacts that are 'closer to home' or more personal in the sense that they may fundamentally impact what it means to be human and interact with our world, we should be aware of this anthropocentric perspective with any kind of technology we use because technology in and of itself is not always good or bad. Thus, my argument can be used and applied to various facets of our interaction with technology.

Although we may still be in the dark as far as some aspects of framing ethical conditions or boundaries go given the novelty of the technology, it is important for us to be prepared for the way in which this technology may change our moral and social lives. If we do not grant this any consideration now, before this type of technology is integrated into society, it may be too late to try to ensure that this technology is beneficial to us. My suggestion for an anthropocentric focus in terms of how we use sexbot technology, the potential moral and social harms that may arise, and possible ethical conditions for human-sexbot interaction we may consider so as to prevent these harms is one way in which we may consider preventing the moral and social harms of the utilisation of sexbot technology in general. As was mentioned above, ensuring beneficial AI in general, and ensuring the beneficiality of sexbot technology, is a big task. However, the arguments I have made add an important contribution to the discussion of beneficial AI in general (given an anthropocentric perspective may be considered in relation to the general utilisation of technology), and the beneficiality of sexbots in particular that will contribute to taking the discussion forward.

Bibliography

Ackerman, E., 2016. *SAM Brings Much-Needed Robotic Assistance to Senior Living Facilities*. [Online] Available at: https://spectrum.ieee.org/automaton/robotics/medical-robots/luvozo-sam-robotic-assistance-to-senior-living-facilities [Accessed 20 June 2020].

Adshade, M., 2017. 'Sexbot-Induced Social Change: An Economic Perspective'. In: J. Danaher & N. McArthur, eds. *Robot Sex: Social and Ethical Implications*. Cambridge, MA: MIT Press, pp. 289-300.

African Union. 2019. *African Digital Transformation Strategy and African Union Communication and Advocacy Strategy Among Major AU Initiatives in Final Declaration of STCCICT3*. [Online] Available at: https://au.int/en/pressreleases/20191026/african-digital-transformation-strategy-and-african-union-communication-and [Accessed 4 August 2020].

AI for Good Global Summit. 2020. *Reports*. [Online] Available at: https://aiforgood.itu.int/reports/ [Accessed 4 August 2020].

AI Now. 2019. *2019 Report*. [Online] Available at: https://ainowinstitute.org/AI_Now_2019_Report.pdf [Accessed 4 August 2020].

Anderson, M. & Anderson, S. L., 2007. 'Machine Ethics: Creating an Ethical Intelligent Agent'. *AI Magazine*, 28(4), pp. 15-26.

Appel, J. M., 2010. 'Sex Rights for the Disabled?'. *Journal of Medical Ethics*, 36, pp. 152-154.

Arnold, T. & Scheutz, M., 2018. 'HRI Ethics and Type-Token Ambiguity: What Kind of Robotic Identity is Most Responsible?'. *Ethics and Information Technology*. [Online] Available at: https://link.springer.com/article/10.1007/s10676-018-9485-1 [Accessed 29 August 2019].

Asaro, P., 2006. 'What Should We Want from a Robot Ethic?'. *International Review of Information Ethics*, 6(12), pp. 9-16.

Asaro, P. M., 2012. 'A Body to Kick, but Still No Soul to Damn: Legal Perspectives'. In: P. Lin, K. Abney & G. A. Bekey, eds. *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge, MA: MIT Press, pp. 169-186.

Avramides, A., 2019. 'Other Minds'. *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), Edward N. Zalta, ed. [Online] Available at: https://plato.stanford.edu/archives/sum2019/entries/other-minds/ [Accessed 4 August 2020].

BAAI. 2019. *Beijing AI Principles*. [Online] Available at: https://www.baai.ac.cn/news/beijing-ai-principles-en.html [Accessed 4 August 2020].

Bartlett, P., 2012. 'Sex, Dementia, Capacity and Care Homes'. *Liverpool Law Review*, 31, pp. 137-154.

Bartlett, K. T., Rhode, D. & Grossman, J. L., 2013. *Gender and Law: Theory, Doctrine, Commentary*. 6 ed. New York: Wolters Kluwer.

BBC News, 2019. *Police Officers Raise Concerns About 'Biased' AI Data*. [Online] Available at: https://www.bbc.com/news/technology-49717378 [Accessed 17 October 2019].

Beck, J., 2013. *Married to a Doll: Why One Man Advocates Synthetic Love*. [Online] Available at: https://www.theatlantic.com/health/archive/2013/09/married-to-a-doll-why-one-man-advocates-synthetic-love/279361/ [Accessed 11 September 2019].

Becker, L. C., 1986. *Reciprocity*. Chicago: University of Chicago Press.

Behdadi, D. & Munthe, C., 2020. 'A Normative Approach to Artificial Moral Agency'. *Minds and Machines*, 30(2), pp. 195-218.

Bekey, A. G., 2012. 'Current Trends in Robotics: Technology and Ethics'. In: P. Lin, K. Abney & G. A. Bekey, eds. *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge, MA: MIT Press, pp. 17-34.

Bentham, J., 1982. In: J.H. Burns and H.L.A. Hart, eds. *An Introduction to the Principles of Morals and Legislation*. London: Methuen.

Berlin, I., 1969. 'Two Concepts of Liberty'. In: *Four Essays on Liberty*. Oxford: Oxford University Press, pp. 118-172.

Birks, D., 2018. 'How Wrong is Paternalism?'. *Journal of Moral Philosophy*, 15, pp. 136-163.

Boden, M. A., Bryson, J. J., Caldwell, D., Dautenhahn, K., Edwards, L., Kember, S., Newman, P., Parry, V., Pegman, G., Rodden, T., Sorell, T., Wallis, M., Whitby, B. &

Winfield, A. F. T., 2017. 'Principles of Robotics: Regulating Robots in the Real World'. *Connection Science,* 29(2), pp. 124–129.

Boltuć, P., 2017. 'Church-Turing Lovers'. In: P. Lin, K. Abney & R. Jenkins, eds. *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*. Oxford University Press, pp. 214-228.

Borenstein, J. & Pearson, Y., 2010. 'Robot Caregivers: Harbingers of Expanded Freedom for All?'. *Ethics and Information Technology*, 12(3), pp. 277-288.

Bostrom, N., 2014. *Superintelligence: Paths, Dangers, Strategies.* New York: Oxford University Press.

Bostrom, N. & Yudkwosky, E., 2014. 'The Ethics of Artificial Intelligence'. In: K. Frankish & W. M. Ramsey, eds. *The Cambdridge Handook of Artificial Intelligence*. Cambridge: Cambridge University Press, pp. 316-334.

Breazeal, C., 2002. *Designing Sociable Robots*. Cambridge, MA: MIT Press.

Breazeal, C., 2003. 'Towards Sociable Robots'. *Robots and Autonomous Systems*, 42, pp. 167-175.

Breazeal, C., 2004. 'Social Interactions in HRI: The Robot View'. *IEEE Transactions on Systems, Man, and Cybernetics*, 34(2), pp. 181-186.

Bringsjord, S. & Govindarajulu, N. S., 2018. 'Artificial Intelligence'. *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), Edward N. Zalta, ed. [Online] Available at: https://plato.stanford.edu/archives/sum2020/entries/artificial-intelligence/ [Accessed 4 August 2020].

Broadbent, E., 2017. 'Interactions with Robots: The Truths We Reveal About Ourselves'. *Annual Review of Psychology*, 68, pp. 627-652.

Brundage, M., 2014. 'Limitations and Risks of Machine Ethics'. *Journal of Experimental & Theoretical Artificial Intelligence*, 26(3), pp. 355-372.

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G. C., Steinhardt, J., Flynn, C., Ó hÉigeartaigh, S., Beard, S., Belfield, H., Farquhar, S., Lyle, C., Crootof, R., Evans, O., Page, M., Bryson, J., Yampolskiy, R. & Amodei, D., 2018. *The Malicious Use of Artificial*

Intelligence: Forecasting, Prevention, and Mitigation. [Online] Available at: https://maliciousaireport.com/ [Accessed 4 August 2020].

Bryson, J. J., 2009. 'Robots Should Be Slaves'. In: Y. Wilks, ed. *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*. Amsterdam: John Benjamins Publishing Company, pp. 63-74.

Bryson, J. J., Diamantis, M. E. & Grant, T. D., 2017. 'Of, For, and By the People: The Legal Lacuna of Synthetic Persons'. *Artificial Intelligence and Law*, 25, pp. 273-291.

Buddy, n.d. *Buddy: The Emotional Robot*. [Online] Available at: https://buddytherobot.com/en/buddy-the-emotional-robot/ [Accessed 20 June 2020].

Calo, R. A., Froomkin, M. & Kerr, I., 2016. *Robot Law*. Northampton, MA: Elgar Publishing.

Carpenter, J., 2017. 'Deus Sex Machina'. In: J. Danaher & N. McArthur, eds. *Robot Sex: Social and Ethical Implications*. Cambrdieg, MA: MIT Press, pp. 261-287.

Carter, I., 2018. 'Positive and Negative Liberty'. *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), Edward N. Zalta, ed. [Online] Available at: https://plato.stanford.edu/archives/win2019/entries/liberty-positive-negative/ [Accessed 4 August 2020].

Chalmers, D., 1995. 'Facing Up to the Problem of Consciousness'. *Journal of Consciousness Studies*, 2(3), pp. 200-219.

Chalmers, D., 2002. *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press.

Chalmers, D., 2010. 'The Singularity'. *Journal of Consciousness Studies*, 17(9-1), pp. 7-65.

Coeckelbergh, M., 2009. 'Personal Robots, Appearance and Human Good: A Methodological Reflection on Roboethics'. *International Journal of Social Robotics*, 13(2), pp. 181-190.

Coeckelbergh, M., 2010a. 'Artificial Companions: Empathy and Vulnerability Mirroring in Human-Robot Relations'. *Studies in Ethics Law and Technology*, 4(3), Article 2.

Coeckelbergh, M., 2010b. 'Health Care, Capabilities, and AI Assistive Technologies'. *Ethical Theory and Moral Practice*, 13, pp. 181-190.

Coeckelbergh, M., 2010c. 'Robot Rights? Towards a Social-Relational Justification of Moral Consideration'. *Ethics and Information Technology*, 12(3), p. 209–221.

Coeckelbergh, M., 2011. 'Moral Responsibility, Technology, and Experiences of the Tragic: From Kierkegaard to Offshore Engineering'. *Science and Engineering Ethics*, 18(1), pp. 35-48.

Computer Ethics Institute, 1992. *Ten Commandments of Computer Ethics.* [Online] Available at: http://computerethicsinstitute.org/publications/tencommandments.html [Accessed 4 August 2020].

Council of Europe. 2020. *CAHAI - Ad Hoc Committee on Artificial Intelligence*. [Online] Available at: https://www.coe.int/en/web/artificial-intelligence/cahai [Accessed 4 August 2020].

Crawford, K. 2017. *The Trouble with Bias. Keynote Address NIPS 2017*. https://www.youtube.com/watch?v=fMym_BKWQzk

Crisp, R., 2011a. *Ethical Concepts and Ethical Theories*. [Online] Available at: https://www.rep.routledge.com/articles/overview/ethics/v-2/sections/ethical-concepts-and-ethical-theories [Accessed 5 May 2019]

Crisp, R., 2011b. *Ethics and Metaethics*. [Online] Available at: https://www.rep.routledge.com/articles/overview/ethics/v-2/sections/ethics-and-metaethics [Accessed 5 May 2019]

Cummings, M. L., 2006. 'Automation and Accountability in Decision Support System Interface Design'. *Journal of Technology Studies*, 32, pp. 23-31.

Damiano, L. & Dumouchel, P., 2018. 'Anthropomorphism in Human–Robot Co-evolution'. *Frontiers in Psychology*, 9, pp. 1-9.

Danaher, J., 2017a. 'Robotic Rape and Robotic Child Sexual Abuse: Should They Be Criminalised?'. *Criminal Law and Philosophy*, 11(1), pp. 71–95.

Danaher, J., 2017b. 'Should We Be Thinking about Robot Sex?'. In: J. Danaher & N. McArthur, eds. *Robot Sex: Social and Ethical Implications*. Cambridge, MA: MIT Press, pp. 3-14.

Danaher, J., 2017c. 'The Symbolic-Consequences Argument in the Sex Robot Debate'. In: J. Danaher & N. McArthur, eds. *Robot Sex: Social and Ethical Implications*. Cambridge, MA: MIT Press, pp. 103-132.

146

Danaher, J., 2018. *Embracing the Robot.* [Online]
Available at: https://aeon.co/essays/programmed-to-love-is-a-human-robot-relationship-wrong [Accessed 17 June 2020].

Danaher, J., 2019. 'The Rise of the Robots and the Crisis of Moral Patiency'. *AI & Society*, 34, pp. 129-136.

Danaher, J., Earp, B. & Sandberg, A., 2017. 'Should We Campaign Against Sex Robots?'. In: J. Danaher & N. McArthur, eds. *Robot Sex: Social and Ethical Implications*. Cambridge, MA: MIT Press, pp. 47-72.

Danaher, J. & McArthur, N., 2017. *Robot Sex: Social and Ethical Implications*. Cambridge, MA: MIT Press.

Darling, K., 2012. *Extending Legal Protection to Social Robots*. [Online] Available at: https://spectrum.ieee.org/automaton/robotics/artificial-intelligence/extending-legal-protection-to-social-robots [Accessed 11 May 2020].

Dautenhahn, K., 1998. 'The Art of Designing Socially Intelligent Agents – Science, Fictions, and the Human in the Loop'. *Applied Artificial Intelligence,* 12, 573-617.

Dautenhahn, K., 2007. 'Socially Intelligent Robots: Dimensions of Human-Robot Interaction'. *Philosophical Transactions of the Royal Society*, 362(1480), pp. 679-704.

Davey, T., 2016. *Who is Responsible for Autonomous Weapons?*. [Online] Available at: https://futureoflife.org/2016/11/21/peter-asaro-autonomous-weapons/ [Accessed 23 April 2020].

DeepMind. *Ethics & Society Principles*. [Online] Available at: https://deepmind.com/about/ethics-and-society [Accessed 4 August 2020].

Deng, B., 2015. 'Machine Ethics: The Robot's Dilemma'. *Nature News,* 523, pp. 24-26.

Dennett, D., 1997. 'Consciousness in Human and Robot Minds'. In: M. Ito, Y. Miyashita & E. T. Rolls, eds. *Cognition, Computation, and Consciousness*. Oxford University Press, pp. 17–29.

Devlin, K., 2018. *Turned On: Science, Sex and Robots*. London: Bloomsbury.

Diamond, M., Jozifkova, E. & Weiss, P., 2010. 'Pornography and Sex Crimes in the Czech Republic'. *Archives of Sexual Behavior,* 40, pp. 1037–1043.

147

Di Nucci, E., 2011. 'Sexual Rights and Disability'. *Journal of Medical Ethics*, 37, pp. 158-161.

Di Nucci, E., 2017. 'Sex Robots and the Rights of the Disabled'. In: J. Danaher & N. McArthur, eds. *Robot Sex: Social and Ethical Implications*. Cambridge, MA: MIT Press, pp. 73-88.

Dickson, E. J., 2018. *Sex Doll Brothels Are Now a Thing. What Will Happen to Real-Life Sex Workers?*. [Online] Available at: https://www.vox.com/the-goods/2018/11/26/18113019/sex-doll-brothels-legal-sex-work [Accessed 26 February 2020].

Duffy, B. R., 2003. 'Anthropomorphism and the Social Robot'. *Robotics and Autonomous Systems*, 42, pp. 177-190.

Dworkin, G., 2020. 'Paternalism'. *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), Edward N. Zalta, ed. [Online] Available at: https://plato.stanford.edu/archives/sum2020/entries/paternalism/ [Accessed 3 May 2020].

Dziergwa, M., Kaczmarek, M., Kaczmarek, P., Kędzierski, J. & Wades-Szydlowska, K., 2018. 'Long-Term Cohabitation with a Social Robot: A Case Study of the Influence of Human Attachment Patterns'. *International Journal of Social Robotics*, 10, pp. 163-176.

Elish, M. C., 2019. 'Moral Crumple Zones Cautionary Tales in Human-Robot Interaction'. *Engaging Science, Technology, and Society,* 5, pp. 40-60.

Ericsson, L. O., 1980. 'Charges Against Prostitution: An Attempt at a Philosophical Assessment'. *Ethics*, 90(3), pp. 335-366.

Ernst, E., Merola, R. & Samaan, D., 2018. *The Economics of Artificial Intelligence: Implications for the Future of Work*, Geneva: International Labour Organisation. [Online] Available at: https://www.researchgate.net/profile/Ekkehard_Ernst/publication/328353684_The_economics_of_artificial_intelligence_Implications_for_the_future_of_work/links/5bc81e60299bf17a1c5b28d2/The-economics-of-artificial-intelligence-Implications-for-the-future-of-work.pdf?origin=publication_detail [Accessed 20 January 2020].

European Commission, 2019. *Building Trust in Human-Centric Artificial Intelligence*, Brussels. [Online] Available at: https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines [Accessed 30 June 2020].

Evans, D., 2010. 'Wanting the Impossible: The Dilemma at the Heart of Human-Robot Relationships'. In: Y. Wilkes, ed. *Close Engagements with Artificial Companions. Key Social, Psychological, Ethical and Design issues*. Amsterdam: John Benjamins Publishing Company, pp. 75-88.

Feenberg, A., 1991. *Critical Theory of Technology*. Oxford: Oxford University Press.

Fehr, B. & Russel, J. A., 1991. 'The Concept of Love Viewed from a Prototype Perspective'. *Journal of Personality and Social Psychology,* 60(3), pp. 425-438.

Floridi, L. & Sanders, J., 2004. 'On the Morality of Artificial Agents'. *Mind and Machines*, 14, pp. 349-379.

Floridi, L., 2013. *The Ethics of Information*. Oxford: Oxford University Press.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Leutge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P. & Vayena, E., 2018. 'AI4People - An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations'. *Minds and Machines*, 28(4), pp. 689-707.

Fong, T., Nourbakhsh, I. & Dautenhahn, K., 2003. 'A Survey of Socially Interactive Robots'. *Robots and Autonomous Systems*, 42, pp. 143-166.

Forrest, K. B., 2019. *The Holographic Judge*. [Online] Available at: https://www.law.com/newyorklawjournal/2019/12/30/the-holographic-judge/ [Accessed 26 February 2020].

Friedman, B., 1997. *Human Values and the Design of Computer Technology*. New York: University of Cambridge Press.

Fritz, N. & Bryant, P., 2017. 'From Orgasms to Spanking: A Content Analysis of the Agentic and Objectifying Sexual Scripts in Feminist, for Women, and Mainstream Pornography'. *Sex Roles*, 77, p. 639–652.

Future of Life Institute, n.d. [Online] Available at: https://futureoflife.org/ [Accessed 4 August 2020].

Future of Life Institute, n.d. *Benefits & Risks of Artificial Intelligence*. [Online] Available at: https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/ [Accessed 21 April 2020].

Future of Life Institute, 2015. *Autonomous Weapons: An Open Letter from AI & Robotics Researchers*. [Online] Available at: https://futureoflife.org/open-letter-autonomous-weapons/ [Accessed 23 April 2020].

Future of Life Institute. 2017. *Asimolar AI Principles*. [Online] Available at: https://futureoflife.org/ai-principles/. [Accessed 19 February 2019].

G7. 2018. *Charlevoix Common Vision for the Future of AI*. [Online] Available at: https://www.international.gc.ca/world-monde/assets/pdfs/international_relations-relations_internationales/g7/2018-06-09-artificial-intelligence-artificielle-en.pdf [Accessed 4 August 2020].

G20. 2019. *G20 Ministerial Statement on Trade and Digital Economy*. [Online] Available at: https://trade.ec.europa.eu/doclib/docs/2019/june/tradoc_157920.pdf [Accessed 4 August 2020].

Goldstein, K., Shem Tov, O. & Prazeres, D., 2018. 'The Right to Privacy in the Digital Age'. In: *Report of the High Commissioner for Human Rights*. United Nations OHCHR. [Online] Available at: https://www.ohchr.org/Documents/Issues/DigitalAge/ReportPrivacyinDigitalAge/PiratePartiesInternational.pdf [Accessed 29 May 2020].

Google. 2018. *Artificial intelligence at Google: Our principles*. [Online] Available at: https://ai.google/principles/ [Accessed 4 August 2020].

Goodrich, M. A. & Schultz, A. C, 2007. 'Human-Robot Interaction: A Survey'. *Foundations and Trends in Human-Computer Interaction,* 1(3), pp. 203-275.

Griggs, T. & Wakabayashi, D., 2018. *How a Self-Driving Uber Killed a Pedestrian in Arizona*. [Online] Available at: https://www.nytimes.com/interactive/2018/03/20/us/self-driving-uber-pedestrian-killed.html [Accessed 11 September 2019].

Gunkel, D. J., 2012. 'Moral Patiency'. In: *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*. Cambridge, MA: MIT Press, pp. 93-157.

Gunkel, D. J., 2017. 'The Other Question: Can and Should Robots Have Rights?'. *Ethics and Information Technology*, 20, pp. 87-99.

Harmonious Artificial Intelligence Principles. 2018. [Online] Available at: http://bii.ia.ac.cn/hai/ [Accessed 4 August 2020].

Hauskellar, M., 2017. 'Automatic Sweethearts for Transhumanists'. In: J. Danaher & N. McArthur, eds. *Robot Sex: Social and Ethical Implications*. Cambridge, MA: MIT Press, pp. 203-218.

Health, Empowerment Rights and Responsibility, 2004. *HERA Action Sheets*, USA: International Women's Rights Coalition.

Hegel, F., Muhl, C., Wrede, B., Hielscher-Fastabend, M. & Sagerer, G., 2009. 'Understanding Social Robots'. In*: Second International Conference on Advances in Computer-Human Interactions (ACHI). IEEE,* pp. 169-174.

High-Level Expert Group on Artificial Intelligence. 2019. *Ethics Guidelines for Trustworthy AI.* [Online] Available at: https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence [Accessed 4 August 2020].

Himma, K. E., 2009. 'Artificial Agency, Consciousness, and the Criteria for Moral Agency: What Properties Must an Artificial Agent Have to be a Moral Agent?'. *Ethics and Information Technology*, 11, pp. 19-29.

Hirschl, R., 2000. '"Negative" Rights vs. "Positive" Entitlements: A Comparative Study of Judicical Interpretations of Rights in an Emerging Neo-Liberal Economic Order'. *Human Rights Quarterly*, 22(4), pp. 1060-1098.

Horstmann, A. C., Block, N., Linhuber, E., Szczuka, J. M., Straßmann, C. & Krämer, N. C., 2018. 'Do a Robot's Social Skills and its Objection Discourage Interactants from Switching the Robot Off?'. *PLOS One*, 13(7). [Online] Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0201581#sec032 [Accessed 29 May 2020].

Hurst, D., 2018. *Japan Lays Groundwork for Boom in Robot Carers*. [Online] Available at: https://www.theguardian.com/world/2018/feb/06/japan-robots-will-care-for-80-of-elderly-by-2020 [Accessed 2 March 2020].

ICRC, 2019. *Autonomy, Artificial Intelligence and Robotics: Technical Aspects of Human Control*, Geneva. [Online] Available at: https://www.icrc.org/en/document/autonomy-artificial-intelligence-and-robotics-technical-aspects-human-control [Accessed 30 May 2020].

IEEE. *Ethically Aligned Design*. 2017. [Online]. Available at:
https://standards.ieee.org/content/dam/ieee-
standards/standards/web/documents/other/ead1e.pdf [Accessed 4 August 2020].

Jacob, P., 2019. 'Intentionality'. *Stanford Encyclopedia of Philosophy* (Winter 2019 Edition),
Edward N. Zalta, ed. [Online] Available at:
https://plato.stanford.edu/archives/win2019/entries/intentionality/ [Accessed 17 May 2020].

Johnson, D. G., 2006. 'Computer Systems: Moral Entities but Not Moral Agents'. *Ethics and
Information Technology*, 8, pp. 195-204.

Jankowiak, W. & Fischer, E. F., 1992. 'A Cross-Cultural Perspective on Romantic Love'.
Ethnology, 31, pp. 149-155.

Jaworska, A. & Tannenbaum, J., 2018. 'The Grounds of Moral Status'. *Stanford
Encyclopedia of Philosophy* (Spring 2018 Edition), Edward N. Zalta, ed. [Online] Available
at: https://plato.stanford.edu/archives/spr2018/entries/grounds-moral-status/ [Accessed 4
August 2020].

Kamm, F. M., 2007. *Intricate Ethics: Rights, Responsibilities, and Permissible Harm*. New
York: Oxford University Press.

Kanda, T., Ishiguro, H., Imai, M. & Ono, T., 2004. 'Development and Evaluation of
Interactive Humanoid Robots'. *Proceedings of the IEEE*, 92(11), pp. 1839-1850.

Kanda, T., Freier, N. G., Severson, R. & Gill, B. T., 2012. '"Robovie, You'll Have to Go into
the Closet Now": Children's Social and Moral Relationships with a Humanoid Robot'.
Developmental Psychology, 48(2), pp. 303-314.

Kirk, R. & Carruthers, P., 1992. 'Consciousness and Concepts'. *Proceedings of the
Aristotelian Society, Supplementary Volumes*, 66, pp. 23-59.

Knapton, S., 2017. *Sex Robots on Way for Elderly and Lonely...but Pleasure-Bots Have a
Dark Side, Warn Experts*. [Online] Available at:
https://www.telegraph.co.uk/science/2017/07/04/sex-robots-way-elderly-lonelybut-pleasure-
bots-have-dark-side/ [Accessed 13 February 2019].

Komatsubara, T., Shiomi, M., Kanda, T., Ishiguro, H. & Hagita, N., 2014. 'Can a Social
Robot Help Children's Understanding of Science in Classrooms?'. *Proceedings of the Second
International Conference on Human–Agent Interaction*, pp. 83–90.

Kranzberg, M., 1986. 'Technology and History: Kranzberg's Laws'. *Technology and Culture*, 27(3), pp. 544-560.

Kurzweil, R., 2005. *The Singularity is Near: When Humans Transcend Biology*. New York: Viking.

Leube, D. T. & Kircher, T. T., 2003. 'Self-Consciousness, Self-Agency, and Schizophrenia'. *Consciousness and Cognition*, 12, pp. 656-669.

Leveson, N. G., 2011. *Engineering a Safer World: Systems Thinking Applied to Safety*. Cambridge, MA: MIT Press.

Levy, D., 2007. *Love and Sex with Robots: The Evolution of Human-Robot Relationships*. New York: Harper Perennial.

Levy, D., 2009. 'The Ethical Treatment of Artificially Conscious Robots'. *International Journal of Social Robotics*, 1(3), pp. 209-216.

Levy, D., 2015. *Robot Prostitutes as Alternatives to Human Sex Workers*. [Online] Available at:
https://www.researchgate.net/publication/237462932_Robot_Prostitutes_as_Alternatives_to_Human_Sex_Workers [Accessed 7 February 2019].

Levy, N., 2002. 'Virtual Child Pornography: The Eroticization of Inequality'. *Ethics and Information Technology*, 4, pp. 319-323.

Liang, Y. L., 1949. 'Reparation for Injuries Suffered in the Service of the United Nations'. *American Journal of International Law*, 43(3), pp. 460-478.

Lin, P., Abney, K. & Bekey, G. A., 2012. *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge, MA: MIT Press.

Lohr, S., 2018. 'It's True: False News Spreads Faster and Wider. And Humans Are to Blame'. *The New York Times*. [Online] Available at:
https://www.nytimes.com/2018/03/08/technology/twitter-fake-news-research.html [Accessed 20 January 2020].

Lumbreras, S., 2017. 'The Limits of Machine Ethics'. *Religions*, 8(100), pp. 2-10.

MacDorman, K. F., 2005. 'Androids as an Experimental Apparatus: Why is There an Uncanny Valley and Can We Exploit It?'. *CogSci-2005Workshop: Toward Social Mechanisms of Android Science*, pp. 106-118.

MacKinnon, C. A., 1992. 'Pornography, Civil Rights, and Speech'. In: C. Itzin, ed. *Pornography: Women, Violence, and Civil Liberties*. New York: Oxford University Press, pp. 456-511.

Matthias, A., 2004. 'The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata'. *Ethics and Information Technology*, 6, pp. 175-183.

McArthur, N., 2017. 'The Case for Sexbots'. In: J. Danaher & N. McArthur, eds. *Robot Sex: Social and Ethical Implications*. Cambridge, MA: MIT Press, pp. 31-45.

McDermott, D., 2008. 'Why Ethics is a High Hurdle for AI', *Indiana: North American Conference on Computers and Philosophy,* Bloomington, Indiana. [Online] Available at: https://pdfs.semanticscholar.org/260b/d7d817d2db2b1a9022b26a5536dd3bc352aa.pdf [Accessed 4 August 2020].

Melson, G. F., Kahn, P. H., Beck, A. & Friedman, B., 2009. 'Robotic Pets in Human Lives: Implications for the Human–Animal Bond and for Human Relationships with Personified Technologies'. *Journal of Social Issues,* 65(3), pp. 545-567.

Migotti, M. & Wyatt, N., 2017. 'On the Very Idea of Sex with Robots'. In: J. Danaher & N. McArthur, eds. *Robot Sex: Social and Ethical Implications*. Cambridge, MA: MIT Press, pp. 15-28.

Mill, J. S., 1986. *On Liberty*. London: Longmans, Green, and Co.

Moon, Y. & Nass, C., 2000. 'Machines and Mindlessness: Social Responses to Computers'. *Journal of Social Issues*, 56, pp. 81-103.

Moor, J., 2006. 'The Nature, Importance, and Difficulty of Machine Ethics'. *IEEE*, 21(4), pp. 18-21.

Nagel, T., 1974. 'What is it Like to be a Bat?'. *Philosophical Review*, 83(October), pp. 435-450.

National Institute of Drug Abuse, 2018. *What Are the Treatments for Heroin Use Disorder?*. [Online] Available at: https://www.drugabuse.gov/publications/research-reports/heroin/what-are-treatments-heroin-use-disorder [Accessed 26 November 2019].

Noddings, N., 1984. *Caring: A Feminine Approach to Ethics and Moral Education*. Berkeley: University of California Press.

Noonan, R., 2003. 'Sex Surrogates: The Continuing Controversy'. In: N. Noonan & R. Francoeur, eds. *Continuum Complete Encyclopedia of Sexuality*. New York: Continuum, pp. 1279-1281.

Nussbaum, M., 2006. *Frontiers of Justice: Disability, Nationality, Species Membership*. Cambridge: Harvard University Press.

Nyholm, S. & Frank, L. E., 2017. 'From Sex Robots to Love Robots'. In: J. Danaher & N. McArthur, eds. *Robot Sex: Social and Ethical Implications.* Cambridge, MA: MIT Press, pp. 219-243.

Nyholm, S. & Frank, L., 2019. 'It Loves Me, It Loves Me Not: Is It Morally Problematic to Design Sex Robots That Appear to Love Their Owners?'. *Techné: Research in Philosophy and Technology*, 23(3), pp. 402-424.

OECD. 2019. *Recommendation of the Council on Artificial Intelligence*. [Online] Available at: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 [Accessed 4 August 2020].

Open Society Foundations, 2019. *Understanding Sex Work in an Open Society*. [Online] Available at: https://www.opensocietyfoundations.org/explainers/understanding-sex-work-open-society [Accessed 2 March 2020].

Paro Robots, 2014. *PARO Therapeutic Robot*. [Online] Available at: http://www.parorobots.com [Accessed 9 July 2019].

Patridge, S. & Jordan, A., 2010. 'Against the Moralistic Fallacy: In Defence of Modest Moralism About Humour'. *Ethical Theory and Moral Practice*, 15(1), pp. 84-95.

Patridge, S., 2010. 'The Incorrigible Social Social Meaning of Video Game Imagery'. *Ethics and Information Technology*, 13(4), pp. 303-312.

Patridge, S., 2013. 'Pornography, Ethics and Video Games'. *Ethics and Information Technology*, 15(1), pp. 25-34.

Perrow, C., 1984. *Normal Accidents: Living with High-Risk Technologies*. New York: Basic Books.

Pierce, C. S., 1998. *The Essential Peirce: Selected Philosophical Writings*. 2 ed. Bloomington: Indiana Press.

Ramey, C. H., 2005. ''For the Sake of Others': The 'Personal' Ethics of Human-Android Interaction'. *Towards Social Mechanisms of Android Science: a COGSCI 2005 Workshop*, Stresa, Italy, 25-26 July, pp. 137-148.

Rani, A., 2013. *The Japanese Men Who Prefer Virtual Girlfriends to Sex.* [Online] Available at: https://www.bbc.com/news/magazine-24614830 [Accessed 17 June 2020].

Riccardi, P., 2010. 'Male Rape: The Silent Victim and the Gender of the Listener'. *The Primary Care Companion to the Journal of Clinical Psychiatry*, 12(6). [Online] Available at: https://www-ncbi-nlm-nih-gov.uplib.idm.oclc.org/pmc/articles/PMC3067991/?report=classic [Accessed 5 March 2020].

Richardson, K., 2015. *The Asymmetrical 'Relationship': Parallels Between Prostitution and the Development of Sex Robots*. [Online] Available at: https://campaignagainstsexrobots.org/the-asymmetrical-relationship-parallels-between-prostitution-and-the-development-of-sex-robots/ [Accessed 5 February 2019].

Richardson, K., 2017. *Sex Robot Matters: Slavery, the Prostituted, and the Rights of Machines.* [Online] Available at: http://technologyandsociety.org/sex-robot-matters-slavery-the-prostituted-and-the-rights-of-machines/ [Accessed 5 February 2019].

Rising, E., 2018. *Why a Sex Robot Brothel Won't Solve the Human Trafficking Problem in the City of Houston*. [Online] Available at: https://www.elijahrising.org/blog/robotbrothels [Accessed 5 March 2020].

Roach, S. M., 2004. 'Sexual Behaviour of Nursing Home Residents: Staff Perceptions and Responses'. *Journal of Advanced Nursing*, 48(4), pp. 371-379.

Robins, B., Dautenhahn, K., Te Boekhorst, R. & Billard, A., 2004. 'Effects of Repeated Exposure to a Humanoid Robot on Children with Autism'. In: S. Keates, P. J. Clarkson, P.

Langdon & P. Robinson, eds. *Designing a More Inclusive World.* London: Springer, pp. 225-236.

Robins, B., 2018. ʻKaspar as a Therapeutic Tool: Encouraging Social Interaction Skills in Children with Autism'. Keynote lecture at the *Social Robotics: The Future of Human-Machine Relationships Research Meet,* Eindhoven University of Technology, 25 January.

Robinson, H., 2017. 'Dualism'. *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta, ed. [Online] Available at: https://plato.stanford.edu/archives/fall2017/entries/dualism/ [Accessed 4 August 2020].

Robohub, 2015. *Robot: Geminoid F.* [Online]
Available at: https://www.nationalgeographic.org/article/robot-geminoid-f/
[Accessed 24 June 2020].

Rosenstand, N., 2000. *The Moral of the Story: An Introduction to Ethics*. 3rd ed. Mayfield Publishing Company.

Royakkers, L. & van Est, R., 2015. 'A Literature Review on New Robotics: Automation from Love to War'. *International Journal of Social Robotics*, 7, pp. 549-570.

Ruddick, S., 1975. 'Better Sex'. In: R. Baker & F. Elliston, eds. *Philosophy and Sex*. Buffalo, NY: Prometheus Books, pp. 83-104.

Russel, S. & Norvig, P., 2009. *Artificial Intelligence: A Modern Approach*. 3 ed. Upper Saddle City, NJ: Prentice Hall.

Russel, S. J., 2019. *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking

Scheutz, M., 2012. 'The Inherent Dangers of Unidirectional Emotional Bonds'. In: P. Lin, K. Abney & G. A. Bekey, eds. *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge, MA: MIT Press, pp. 205-221.

Scheutz, M. & Arnold, T., 2016. *Are We Ready for Sex Robots?*. [Online] Available at: https://www.researchgate.net/publication/301912956_Are_we_ready_for_sex_robots [Accessed 5 February 2019].

Scoccia, D., 2008. 'In Defense of Hard Paternalism'. *Law and Philosophy*, 27, pp. 351-381.

Scoglio, A. A., Reilly, E. D., Gorman, J. A. & Drebing, C. E., 2019. 'Use of Social Robots in Mental Health and Well-Being Research: Systematic Review'. *Journal of Medical Internet Research,* 21(7). [Online] Available at: https://www.jmir.org/2019/7/e13322/ [Accessed 29 May 2020].

Select Committee on Artificial Intelligence, 2018. *AI in the UK: Are We Ready, Willing and Able?*, House of Lords. [Online] Available at: https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf [Accessed 30 June 2020].

Shalev-Shwartz, S. & Ben-David, S., 2014. *Understanding Machine Learning: From Theory to Algorithms*. New York: Cambridge University Press.

Sharkey, A. J. & Sharkey, N., 2010. 'Granny and the Robots: Ethical Issues in Robot Care for the Elderly'. *Ethics and Information Technology*, 14(1), pp. 27-40.

Sharkey, A. J. C., 2016. 'Should We Welcome Robot Teachers?'. *Ethics and Information Technology*, 18, pp. 283–297.

Sharkey, N., van Wynsberghe, A., Robbins, S., Hancock, E., 2017. *Our Sexual Future with Robots: A Foundation for Responsible Robotics Consultation Report.* [Online] Available at: https://responsible-robotics-myxf6pn3xr.netdna-ssl.com/wp-content/uploads/2017/11/FRR-Consultation-Report-Our-Sexual-Future-with-robots-.pdf [Accessed 5 February 2019].

Sheridan, T. B. & Parasuraman, R., 2005. 'Human-Automation Interaction'. *Review of Human Factors and Ergonomics*, 1(1), pp. 89-129.

Singer, P., 1977. *Animal Liberation*. London: Granada.

Skorupski, J., 1998. *Morality and Ethics*. [Online] Available at: https://www.rep.routledge.com/articles/thematic/morality-and-ethics/v-1 [Accessed 20 May 2019].

Smith, B., 1928. 'Legal Personality'. *Yale Law Journal*, 37(3), pp. 283-299.

Snell, J. C., 1997. 'Impacts of Robotic Sex'. *The Futurist,* July-August, p. 32.

SoftBank Robotics, n.d. *Pepper*. [Online] Available at: https://www.softbankrobotics.com/emea/en/pepper [Accessed 20 June 2020].

Sony, 2020. *Aibo*. [Online] Available at: https://us.aibo.com/ [Accessed 20 June 2020].

Sparrow, R., 2004. 'The Turing Triage Test'. *Ethics and Information Technology,* 6(4), pp. 203-213.

Sparrow, R. & Sparrow, L., 2006. 'In the Hands of Machines? The Future of Aged Care'. *Minds and Machines*, 16, pp. 141-161.

Sparrow, R., 2011. 'Robotic Weapons and the Future of War'. In: P. Tripodi & J. Wolfendale, eds. *New Wars and New Soliders: Military Ethics in the Contemporary World*. Surrey, England: Ashgate Publishing Limited, pp. 117-136.

Sparrow, R., 2017. 'Robots, Rape, and Representation'. *International Journal for Social Robotics*, 9(3), pp. 465-477.

Stanford University. 2016. *One Hundred Year Study on Artificial Intelligence (AI100)*. [Online] Available at: https://ai100.stanford.edu/ [Accessed 4 August 2020].

Strikwerda, L., 2011. 'Virtual Child Pornography: Why Images Do Harm from a Moral Perspective'. In: C. Ess & M. Thorseth, eds. *Trust and Virtual Worlds: Contemporary Perspectives*. New York: Peter Lang Publishing, pp. 139-161.

Strikwerda, L., 2017. 'Legal and Moral Implications of Child Sex Robots'. In: J. Danaher & N. McArthur, eds. *Robot Sex: Social and Ethical Implications*. Cambridge, MA: MIT Press, pp. 133-151.

Sullins, J. P., 2006. 'When Is a Robot a Moral Agent?'. *International Review of Information Ethics*, 6(12), pp. 23-30.

Sullins, J. P., 2012. 'Robots, Love and Sex: The Ethics of Building a Love Machine'. *IEEE Transactions on Affective Computing*, 3(4), pp. 398-409.

Tarzia, L., Fetherstonhaugh, D. & Bauer, M., 2012. 'Dementia, Sexuality and Consent in Residential Aged Care Facilities'. *Journal of Medical Ethics*, 38, pp. 609-613.

Tavani, H. T., 2018. 'Can Social Robots Qualify for Moral Consideration? Reframing the Question about Robot Rights'. *Information*, 9(73), pp. 1-16. [Online] Available at: https://www.mdpi.com/2078-2489/9/4/73/htm [Accessed 13 April 2020].

Taylor, L., 2017. *Sex Robots: Perverted or Practical in Fight Against Sex Trafficking?*. [Online] Available at: https://www.reuters.com/article/science-robots-sextrafficking/sex-

robots-perverted-or-practical-in-fight-against-sex-trafficking-idUSL1N1KB0M0 [Accessed 5 March 2020].

Tegmark, M., 2018. *Life 3.0*. Penguin Books.

The Human Rights Council. 2019. *The Right to Privacy in the Digital Age*. [Online] Available at: https://digitallibrary.un.org/record/3837297?ln=en#record-files-collapse-header [Accessed 4 August 2020].

The School of ICASA, n.d. *Sexual Surrogate Partner Therapy*. [Online] Available at: https://www.icasa.co.uk/surrogate-partner-therapy [Accessed 26 November 2019].

The United States Department of Justice Archives, 2012. *An Updated Definition of Rape*. [Online] Available at: https://www.justice.gov/archives/opa/blog/updated-definition-rape [Accessed 4 March 2020].

Thomason, R., 2016. 'Logic and Artificial Intelligence'. *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), Edward N. Zalta, ed. [Online] Available at: https://plato.stanford.edu/archives/sum2020/entries/logic-ai/ [Accessed 1 May 2019].

Torrance, S., 2008. 'Ethics and Consciousness in Artificial Agents'. *AI & Society,* 22, pp. 495–521.

Torrance, S., 2013. 'Artificial Agents and the Expanding Ethical Circle'. *AI & Society*, 28, pp. 399-414.

Turkle, S., 1995. *Life on Screen*. New York: Simon and Schuster.

Turkle, S., 2005. *The Second Self: Computers and the Human Spirit*. Cambridge, MA: MIT Press.

Turkle, S., 2006. 'A Nascent Robotics Culture: New Complicities for Companionship'. *AAAI Technical Report Series*. [Online] Available at: http://web.mit.edu/sturkle/OldFiles/www/nascentroboticsculture.pdf [Accessed 14 November 2019].

Turkle, S., 2007. 'Authenticity in the Age of Digital Companions'. *Interaction Studies*, 8(3), pp. 501-517.

Turner, J., 2019. 'Why Robot Rights?'. In: *Robot Rules: Regulating Artificial Intelligence*. Cham: Palgrave Macmillan, pp. 145-171.

UNESCO. 2017. *Recommendation on Science and Scientific Researchers*. [Online] Available at:http://portal.unesco.org/en/ev.php-URL_ID=49455&URL_DO=DO_TOPIC&URL_SECTION=201.html [Accessed 4 August 2020].

UNESCO. 2019. *Internet Universality Indicators: A Framework for Assessing Internet Development.* [Online] Available at: https://unesdoc.unesco.org/ark:/48223/pf0000367617?posInSet=1&queryId=a46642b0-1893-4f04-9bfb-b832b0851652 [Accessed 4 August 2020].

UNI Global Union. 2017. *Top 10 Principles for Ethical Artificial Intelligence*. [Online] Available at: http://www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf [Accessed 4 August 2020].

UN Secretary-General's High-level Panel on Digital Cooperation. 2019. *The Age of Digital Interdependence.* [Online] Available at: https://www.un.org/en/pdfs/HLP%20on%20Digital%20Cooperation%20Report%20Executive%20Summary%20-%20ENG.pdf [Accessed 4 August 2020].

University of Montréal. 2018. *Montréal Declaration for a Responsible Development of AI*. [Online] Available at: https://www.montrealdeclaration-responsibleai.com/the-declaration [Accessed 4 August 2020].

Vallor, S., 2011. 'Carebots and Caregivers: Sustaining the Ethical Ideal of Care in the Twenty-First Century'. *Philosophy & Technology*, 24, p. 251–268.

Vallor, S., 2015. 'Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character'. *Philosophy and Technology,* 28, pp. 107-124.

Vannoy, R., 1980. *Sex Without Love: A Philosophical Exploration*. Amherst, NY: Prometheus Books.

van Wynsberghe, A. & Robbins, S., 2019. 'Critiquing the Reasons for Making Artificial Moral Agents'. *Science and Engineering Ethics,* 25(3), pp. 719-735.

Vaughan, D., 1996. *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA*. Chicago, IL: University of Chicago Press.

Verbeek, P. P., 2011. *Moralizing Technology: Understanding and Designing the Morality of Things.* Chicago: University of Chicago Press.

Wagner, C., 2018. 'Sexbots: The Ethical Ramifications of Social Robotics' Darkside'. *AI Matters,* 3(4), pp. 52-58.

Wallach, W. & Allen, C., 2009. *Moral Machines: Teaching Robots Right from Wrong*. Oxford: Oxford University Press.

Wang, W. & Siau, K., 2018. 'Ethical and Moral Issues with AI: A Case Study on Healthcare Robots', *Twenty-fourth Americas Conference on Information Systems,* New Orleans, August 2018. [Online] Available at:
https://www.researchgate.net/publication/325934375_Ethical_and_Moral_Issues_with_AI
[Accessed 23 June 2020].

Waytz, A., Cacioppo, J. & Epley, N., 2014. 'Who Sees Human? The Stability and Importance of Individual Difference in Anthropomorphism'. *Perspectives on Psychological Science*, 5, pp. 219-232.

Weitzer, R., 2012. *Legalizing Prostitution: From Illicit Vice to Lawful Business*. New York: New York University Press.

Wenar, L., 2020. 'Rights'. *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), Edward N. Zalta, ed. [Online] Available at:
https://plato.stanford.edu/archives/spr2020/entries/rights/ [Accessed 4 August 2020].

Whelan, S., Murphy, K., Barrett, E., Krusche, C., Santorelli, A. & Casey, D., 2018. 'Factors Affecting the Acceptability of Social Robots by Older Adults Including People with Dementia or Cognitive Impairment: A Literature Review'. *International Journal of Social Robotics*, 10(5), pp. 643-668.

Whitby, B., 2012. 'Do You Want a Robot Lover? The Ethics of Caring Technologies'. In: P. Lin, K. Abney & G. A. Bekey, eds. *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge, MA: MIT Press, pp. 233-248.

Wollstonecroft, M., 2009. *A Vindication of the Rights of Woman.* New York: Norton & Company.

World Economic Forum. 2019. *Global Technology Governance: A Multistakeholder Approach*. [Online] Available at:
http://www3.weforum.org/docs/WEF_Global_Technology_Governance.pdf [Accessed 4 August 2020].

Zahavi, D. & Parnas, J., 1998. 'Phenomenal Consciousness and Self-Awareness: A Phenomenological Critique of Representational Theory'. *Journal of Consciousness Studies*, 5(5-6), pp. 687–705.