

Rethinking Privacy in the Age of Social Robots

Coggins, T.N.

DOI

[10.4233/uuid:bf291028-ef7e-4a6c-9a22-9bae9fde591c](https://doi.org/10.4233/uuid:bf291028-ef7e-4a6c-9a22-9bae9fde591c)

Publication date

2024

Document Version

Final published version

Citation (APA)

Coggins, T. N. (2024). *Rethinking Privacy in the Age of Social Robots*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:bf291028-ef7e-4a6c-9a22-9bae9fde591c>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Rethinking Privacy in the Age of Social Robots

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Monday 4 March at 12.30 o'clock

by

Thomasin Nathaniel COGGINS
Master of Arts in Philosophy (research), University of Amsterdam, the Netherlands
born in Newport, Isle of Wight, the United Kingdom

This dissertation has been approved by the promotor.

Composition of the doctoral committee:

Rector Magnificus,	Delft University of Technology, chairperson
Prof. dr. ir. I.R. van de Poel	Delft University of Technology, promotor
Dr. O. Kudina	Delft University of Technology, copromotor

Independent members:

Prof. dr. ir. D.A. Abbink	Delft University of Technology
Prof. dr. W.F.G. Haselager	Radboud University
Prof. dr. S. Nyholm	Ludwig Maximilian University of Munich
Dr. M.H. Nagenborg	University of Twente
Prof. dr. S. Roeser	Delft University of Technology (reserve member)

This research was funded by H2020 European Research Council, grant number: 788321.

Keywords: privacy, social robots, human-robot-interactions, housework, norms, performativity, robot ethics

Printed by: The American Book Center

Cover design: Menah Wellen

Copyright: © 2024 by Thomasin Nathaniel Coggins

Table of Contents

1 Introduction	6
<i>1.1. Robot ethics and privacy</i>	6
<i>1.2. Robots and unwelcome social interactions</i>	12
<i>1.3. Privacy as social disengagement</i>	16
<i>1.4. Chapter overview</i>	21
2 Called Back Onstage: Dramaturgic Analysis, Domestic Social Robots, and Privacy	27
<i>2.1. Introduction</i>	27
<i>2.2. Dramaturgic analysis</i>	28
<i>2.3. Being called back onstage by robots</i>	31
<i>2.4. Privacy as being offstage</i>	34
<i>2.5. Conclusion</i>	37
3 Friends Are Not “Electric” (Characters): A Sociological Case Against Human-Robot Friendships	39
<i>3.1. Introduction</i>	39
<i>3.2. The performance account</i>	40
<i>3.3. What is passing?</i>	44
<i>3.4. Friends are not “electric” (characters)</i>	48
<i>3.5. Privacy issues</i>	52
<i>3.6. Conclusion</i>	55
4 More Work for Roomba? Domestic Robots, Housework, and the Production of Privacy	57
<i>4.1. Introduction</i>	57
<i>4.2. Housework and the production of privacy</i>	59
<i>4.3. A brief critique of domestic technologies</i>	62

4.4. <i>Domestic robots and housework</i>	66
4.4. <i>Conclusion</i>	73
5 Maintaining One’s Home, Maintaining Oneself: A Critical Introduction to the Philosophy of the Home	75
5.1. <i>Introduction</i>	75
5.2. <i>The relief and production accounts</i>	76
5.3. <i>A detour through “maintenance”</i>	82
5.4. <i>The maintenance account</i>	86
5.5. <i>Conclusion</i>	93
6 The Seven Troubles with Norm-Compliant Robots	95
6.1. <i>Introduction</i>	95
6.2. <i>What are norms?</i>	97
6.3. <i>Seven troubles with norms</i>	99
6.3.1. <i>Norm biases</i>	100
6.3.2. <i>Paternalism</i>	101
6.3.3. <i>Tyranny of the majority</i>	104
6.3.4. <i>Pluralistic ignorance</i>	106
6.3.5. <i>Paths of least resistance</i>	109
6.3.6. <i>Outdated norms</i>	112
6.3.7. <i>Robot-induced norm change</i>	115
6.4. <i>Conclusion</i>	119
8 Conclusion	122
8.1. <i>Is hell other people (and now robots)?</i>	122
8.2. <i>Recommendations</i>	124
8.3. <i>Concluding remarks</i>	129

9 Bibliography	131
10 Summary	150
11 Samenvatting	154
11 Acknowledgements	158
12 About the Author	159
13 List of Publications	160

1 Introduction

1.1. Robot ethics and privacy

The Soul selects her own Society —
Then — shuts the Door —
To her divine Majority —
Present no more —

Unmoved — she notes the Chariots — pausing —
At her low Gate —
Unmoved — an Emperor be kneeling
Upon her Mat —

I've known her — from an ample nation —
Choose One —
Then — close the Valves of her attention —
Like Stone —

Emily Dickinson, *The Soul Selects Her Own Society*

In this thesis, I aim to inspire other robot ethicists and scholars from adjacent fields to think differently about privacy. I have spent the last four years researching how robots, mostly installed inside people's homes, affect their users' privacy. While doing so, I read many outstanding academic contributions dedicated to this topic. Indeed, privacy stands among the most discussed values in the robot ethics literature (Kahn et al., 2006; Leenes et al, 2017; Coeckelbergh, 2020; Pasquale, 2020). However, robot ethicists usually conceptualize privacy in a way that limits the scope of their evaluations. Many, if not most, robot ethicists who discuss privacy in their work imply or explicitly claim that this value amounts to the appropriate distribution of information. I contend that we would fail to recognize significant privacy concerns raised by robots as privacy concerns at all if we accepted this dominant conceptualization. As I will argue in this introduction, losing one's privacy does not necessarily mean losing control over one's information. For instance, feeling as though we must interact with someone when we wish or need solitude would prevent us from enjoying

our privacy. This person's unwelcome presence would cause us distress without necessarily robbing us of information we deem private. I argue throughout this thesis that robots that simulate what it is like to interact with another person can produce privacy concerns of this kind too.

As far as I know, this topic has not yet been systematically investigated. Hence, the aim of this thesis is to offer an account of privacy, unlike those usually found in the robot ethics literature. This account will be developed by interpreting social, political, and philosophical theory that recognizes that we need to sporadically distance ourselves from the people we coordinate with in public to live good lives. We generally call this practice "having privacy". If a robot prevents its users from realizing the need mentioned above, I argue, we should say it jeopardizes this person's privacy - even when this machine does not negatively affect their ability to choose with whom they share information. This is the central claim of my thesis. I will explain why I decided to defend this claim in more detail later in this introduction. Before I do, however, I will provide a review of how robot ethicists tend to interpret privacy to draw attention to the theoretical shortfall I aim to alleviate in this thesis.

I began my PhD a few months after my promoter, prof. Ibo van de Poel published an article in the *Journal of Ethics and Information Technology* titled *Design for Value Change* (van de Poel, 2018). Van de Poel suggests in this contribution that the ethics of technology, in general, has more-or-less settled on a definition of privacy (van de Poel, 2018). Nowadays, scholars from this field tend to treat privacy as informational privacy; therefore, they rarely acknowledge alternative interpretations of this value. This interpretive bias makes sense when we consider the innovations that have shaped the last few decades of human history (de Wildt, van de Poel, & Chappin, 2022).

We often call the times we live in the Information Age due to our reliance on information communication technologies (Castells, 1996). Many of us could not do our jobs, socialize with friends, or buy groceries without using these technologies (Turow, 2018). We depend on them. However, we know that doing so can, and often does, compromise our privacy (Barnes, 2006). Controlling who knows what about us has become exceptionally difficult. Indeed, we lose some control over our information whenever we scroll through our social media feeds or send an unencrypted email to our colleagues. When other actors gain access to this information, they can cause us various privacy harms (Solove, 2006). For instance, hackers

can steal our identities by piecing together revealing data we inadvertently disclosed online. And perhaps more worryingly, governmental agencies and corporate actors can monitor us via the data footprints we leave whenever we interact with our mobile phones or personal computers (Bouk, 2017; Zubuff, 2019)

Thanks to the work of countless academic researchers, legal practitioners, and members of non-government organizations, there are well-established methods available to identify and address informational privacy issues like the ones I just described. For instance, ethically and politically informed engineering strategies established over the past two decades, such as Value Sensitive Design, Privacy by Design, and Data Feminism, have successfully shown that organizations can create technologies that preserve people's privacy when they follow design principles devised for this purpose (Friedman et al., 2006; Cavoukian, 2011; D'Ignazio and Klein; 2020). Additionally, European citizens now have legal rights that protect them from invasive data collection and processing practices due to the implementation of the General Data Protection Regulation in 2019 (European Council, 2019).

I applaud these and the many other attempts to rectify the privacy issues brought about by the Information Age. Losing control of our information often exposes us to harm we do not deserve to face – including the ones I outlined two paragraphs ago (van de Hoven, 1998). However, as van de Poel suggests, these efforts chiefly focus on improving one aspect of our privacy, namely our informational privacy. I contend that robot ethics has inherited this way of thinking from other disciplines that seek to evaluate information communication technologies' impact on privacy. I will now evidence this claim by using other robot ethicists' work on privacy to reconstruct how, I posit, researchers from this field usually assess whether a robot will jeopardize its users' privacy.

Suppose a company creates a humanoid robot that provides companionship to its users inside their homes. Several robots like this have reached the market over the past decade, such as Jibo, Buddy and Miko (Ulanoff, 2017; Milliez, 2018; Wairagkar et al. 2020). This hypothetical machine relies on a collection of sensors to gather input data, including two cameras that represent its eyes and an internal microphone. Furthermore, it can "speak". It constantly interfaces with a natural language processing program stored in its manufacturer's server to create the impression that it understands and can respond to its users' utterances. When someone walks past the robot, it greets them and initiates friendly small talk that can

lead to lengthier, more personalized conversations. These features should raise alarm bells because they lay the grounds for a host of informational privacy concerns.

First and foremost, the robot would record its users at homes (Urquhart, Reedman-Flint & Leesakul, 2019). Therefore, its users may say or do things around it that they do not want it to record. For instance, the robot might overhear a couple discussing their financial situation (Lutz, Schöttler & Hoffman, 2019), or observe someone who believes that they are alone getting changed (Lutz and Tamo-Larriueux, 2021). Additionally, its ability to simulate what it is like to chat with a friendly person may encourage its users to disclose information they do not want anyone else to hear. Over time, users may start believing that they can trust this robot and tell it about their relationships, interests, and opinions (Sharkey & Sharkey, 2010). We usually see the information I just described as private for good reasons. If someone learnt about our finances, recorded us in a state of undress, or discovered what we tell our friends in confidence, they could harm us.

The robot almost certainly will transmit this information to its manufacturers. They likely need these recordings to improve their products. For instance, natural language processing programs “learn” how to sound more human-like by analyzing (ideally) an ever-growing corpus of audial recordings (Kurana, Khatter, & Singh, 2023). Likewise, technology companies often employ people to review videos recorded by their products to manually train their computer vision algorithms (Klaus-Scheuermann, Hanna, & Denton 2021). Furthermore, its manufacturer might sell this data to other parties. And hackers could acquire it by exploiting the robot’s cybersecurity weaknesses (Dennings et, al. 2009). Hence, users may lose control of their information by interacting with this machine. Especially if they signed an end-user agreement that stipulates that the robot’s manufacturer may use the data it collects as they wish – which is often the case (Solove, 2012; Terpstra et al. 2019).

The robot increases the likelihood its users will experience informational privacy harms that could damage their reputations, cause them emotional distress, or injure them financially (Solove, 2008; Warren & Brandies, 1890). For instance, a hacker could blackmail them by threatening to publish the information the robot gathered online unless they pay a ransom. And powerful institutional actors, such as insurance agencies or governmental bureaus, could use this data to profile users and evaluate whether they deserve certain goods or services based on this categorization (O’Neil, 2018). Nefarious actors could harm the robot’s users in many

other ways after acquiring sensitive data collected by this machine. For brevity's sake, I will not list these additional informational privacy harms here. Instead, I will assess the claim I made a few paragraphs ago.

If we accept that robot ethicists tend to evaluate whether robots will jeopardize their users' privacy in the manner recounted above, what is left unsaid? I contend that this prevalent mode of analysis overlooks critical, interrelated insights found within the sociologically focused literature on robots and privacy concerning people's experiences of human-to-human and human-to-robot interactions.

Robots have capabilities that other information communication technologies typically do not possess. Research shows that people tend to treat robots that seem to know how to decode and encode socially meaningful messages as though they can (Graaf & Malle, 2017). When robots "speak" or otherwise signal that they understand verbal or non-verbal forms of communication, they often prompt their users to rely on their social know-how when interacting with these machines (Breazeal, 2006; Calo, 2014; Nyholm, 2019). For instance, the hypothetical companionship robot I described earlier encourages its users to chat with it by mimicking how friendly people speak and behave. It simulates what it is like to talk with a friend, and it can only accomplish this feat after it convinces its users to act as though it can understand and respond to their utterances like a human plausibly would. I will argue, throughout this thesis, that robots' ability to trigger social responses from their users produces privacy concerns that do not involve the collection, transmission, or misuse of information.

Whereas the ethical literature on information communication technologies, including robots, generally portrays privacy as informational privacy, privacy scholars generally maintain that this value represents significantly more than the appropriate distribution of information. Some examples will help illustrate this point. If a stranger asked us unwelcome questions about our family histories, which we refused to answer, we may feel as though they intruded on our privacy (Waldman, 2018). Likewise, when a house guest overstays their welcome and does not realize that we want them to leave, we might claim they have disrespected our privacy too (Inness, 1994, Solove, 2008). These cases, and others like them, demonstrate that we can experience privacy harms even when nobody learns anything new about us. Indeed, they show that we become distressed when people fail to appreciate our unwillingness to interact with

them, thus implying that we value when they recognize that we want to be left alone. I will provide more evidence for this claim in section 2 of this introduction.

If a robot behaves like the stranger or houseguest described above, I posit it will produce similar privacy concerns to those created by these two figures. As I suggested earlier, people often relate to robots as though they were humans. Therefore, we can assume that people will experience distress when a robot fails to respect their unwillingness to interact with it. Even if a robot that behaved this way did not gather information from its users, it would still produce privacy concerns of this kind. It jeopardizes its user privacy by engaging with them when they want or need to be left alone (Warren & Brandies, 1890) rather than increasing the likelihood that they will lose control of their private information. If my argument holds, it illustrates that we cannot exclusively rely on conceptualizations of privacy that treat this value as the appropriate distribution of information to identify or address the privacy concerns raised by robots - because these machines produce privacy concerns beyond this conceptualizations' purview.

I spend approximately two-thirds of this thesis (mostly in Chapters 2, 3 and 4) developing theoretical insights that enable us to assess whether a robot will jeopardize its users' privacy in the manner sketched above. Specifically, I argue that many robots will make it harder for their users to adequately avoid or set aside social engagements because they initiate them, often in contexts where people otherwise expect privacy. I conducted highly interdisciplinary, conceptual research to produce these chapters, and this thesis, in general. Most notably, my work draws heavily from sociology and political theory that highlights that humans cannot successfully interact with one another without feeling assured that they (or someone else) can eventually end said interaction without creating social backlash. Many prominent privacy scholars have used a similar strategy to explain privacy's normative value since privacy scholarship became a fully-fledged sub-field of ethics at the end of the 19th century (Glancy, 1979 Westin, 1964; Rössler, 2004). Nonetheless, I believe I among the first scholars to use this approach to evaluate contemporary robots' impact on their users' privacy.

The rest of this introduction proceeds as follows. In Section 1.2., I present a literature review of contributions from robot ethics that implicitly or explicitly address the notion that robots will intrude upon their user's privacy by mimicking human social behavior. In Section 1.3., I introduce concepts and observations from privacy scholarship that I frequently reference

throughout this thesis concerning our need to disengage from social interactions. And in Section 1.4., I explain the thesis' thematic structure and outline the contents of its chapters.

1.2. Robots and unwelcome social interactions

I am not the first scholar to suggest that robots may disturb people by simulating social interactions. However, the robot ethics literature contains very few contributions that interpret this phenomenon via privacy scholarship that recognize that we need and deserve relief from social interactions. In this section, I will provide a review of similar work to my own that shows that robots can dissuade people from performing tasks or motivate them to assume new responsibilities by creating the impression that they understand human social behavior. These articles and books imply that we must sporadically withdraw from social engagements to live good lives without clarifying precisely why we value this practice. As such, I will explain social withdrawal's ethical import by interpreting several prominent privacy scholars' research on this subject in section 1.3. of this introduction. This discussion will serve as the theoretical basis of my thesis.

Journalist, Joel Garreau, was perhaps the first person to document the type of ethical concern I will discuss in this thesis. In 2006, Garreau published a report in the Washington Post on the state-of-art in robotics that included an anecdote concerning several MIT graduate students' experiences sharing a workspace with KISMET: a robot designed to replicate facial expressions humans use to communicate non-verbally with one another (Garreau, 2006). According to Garreau, the students would place a curtain between themselves and KISMET during late nights research sessions because the robot's gaze distracted them from their work. Making eye contact with KISMET would prompt the robot to begin an interaction and the students did not want to engage with KISMET while concentrating on other tasks. Being a reporter, Gareau recounts this story without providing much commentary. Nonetheless, several robot ethicists have cited this report in their work chiefly to demonstrate that people often behave as though a human were present when they encounter robots like KISMET (Scheutz, 2014; Geertz, 2018).

Sherry Turkle, a psychologist who studied people's interactions with KISMET (Turkle, 2007), highlights throughout her 2011 book *Alone Together* that robots that appear to have social needs can make their users feel like they must interact with them (Turkle, 2011). Most notably, she claims that people often act as though robots which resemble infants need care and

affection because these machines seem vulnerable. We know we must constantly dote on babies to ensure they do not become upset. Turkle posits that people often treat infant-like robots this way and modify their behavior to accommodate these machines' non-existent social needs. When this happens, she argues, people may begin putting the robot's imagined needs before their own (Turkle, 2011, p. 75, 103-127). They may find it difficult to leave a robot alone because they believe it needs their attention, thus decreasing the amount of time they can dedicate to tasks that require them to stop interacting with this machine (e.g., relaxation or socializing with their friends and family) (Turkle, 2011, p.75, 1-3-127). Turkle suggests that other robots that seem to have social wants and needs will produce similar effects. A robot that suggests it wishes to interact with someone may make its users feel like they should and persuade them to drop whatever they were doing to attend to it (Turkle, 2011, p. 37-41).

Whereas Garreau and Turkle chiefly discusses how robots can dissuade their users from focusing on tasks that demand their attention, other researchers have highlighted that robots that mimic human social behavior may cause people to assume new responsibilities. For instance, Anna Dobrosovstnova and Glenda Hannibal argue that robots designed to serve as stand-ins for human service workers may increase the amount of emotional labor people working alongside them must perform (Dobrosovstnova & Hannibal, 2020). We expect people who work in the service industry to appear friendly and helpful when we interact with them. They create the impression that they care about their clients' and customers' experiences, even when this is not the case (Urry, 2008). Sociologists call this practice "emotional labor" (Hochschild, 1974).

Whereas human service workers cannot always relate to people this way (e.g., they take breaks and often stop performing emotional labor to complete other workplace tasks), robots can. For instance, a robotic shop assistant could appear perpetually eager to please. Dobrosovstnova and Hannibal claim that human service workers who work alongside a machine like this may start to emulate its computer-generated emotional state to ensure people think they are as professionally competent as the robot; potentially leading to "an increased sense of self-alienation and work-related stress." (Dobrosovstnova & Hannibal, 2020). If this were to happen, the robot in question would effectively compel human staff members to retrain themselves to keep up with or compete with this machine's inhuman capacity to appear amiable, therefore leaving them with less time and energy to attend to workplace responsibilities they value.

Olya Kudina suggests that robots that “talk” may encourage users to develop unwelcome habits that require effort and vigilance to correct (Kudina, 2021). Children, for instance, may pick up behaviors that their parents dislike by interacting with robots that respond to spoken language. Kudina explains that even though children often understand that they are speaking to a technological device rather than a person, they may be less equipped than the adults to resist the conversation patterns and norms that technologies equipped with voice-recognition software promote. Due to technical limitations, speech-recognition software often has trouble processing utterances aside from clear, direct statements (Kudina, 2021). Therefore, children who regularly interact with technologies that simulate conversations may develop speech patterns that sound rude or inappropriate when used to converse with humans.

Kudina reports that parents who let their children interact with technologies of this kind often monitor their offspring’s linguistic development to ensure they know when they should and should not speak as though they were talking to a machine (Kudina, 2021). As such, we can assume that these parents spend more time teaching their children communication skills than they would if they did not own a technology that disrupts this learning process. Adults can develop similar habits too. Research shows that adults who regularly interact with virtual assistants often accidentally use machine-readable language when talking to other people (Smith, 2019). It follows that using said technology inadvertently produces social responsibilities (e.g., the need to rectify undesirable speech patterns) to address a problem the technology itself helped to create by simulating social interactions. Teaching children and oneself how to switch between the different modes of speaking mentioned above would presumably require effort that one could dedicate to more enjoyable or constructive activities.

The contributions outlined above do not explicitly assess the issues they identify as though they were privacy concerns. It is uncommon for robot ethicists to evaluate disruptive interactions produced by robots this way. One notable exception is Ryan Calo, who argues that robots that seem to hold opinions may chill their users’ behavior and, therefore, compromise their privacy in a manner that the conceptualizations of privacy I described in section 1.1. of this introduction cannot address.

When we interact with people who have power over us, for instance, police officers or our managers, we often avoid acting in ways that may displease them. When we see a police officer on a street corner, we may pay more attention to laws and norms that we think they

want us to follow. Likewise, during meetings with our bosses, we usually refrain from discussing subjects that could make them believe we are lazy, unprofessional, or incompetent. Privacy scholars call this phenomenon “a chilling effect” (Solove, 2008; Penney, 2021). Chilling effects are not always problematic; indeed, we often must endure them during interactions with people who we know could make our lives more difficult if we upset them. We cannot live among other people without experiencing this occasionally. However, if someone surveyed us when we deserve to act as we please (e.g., at home), they may harm us by chilling our behavior (Solove, 2008).

Chilling effects cause harm by dissuading individuals from expressing themselves. When someone chills our behavior, they discourage us from speaking freely, acting with less concern for social norms, and generally enjoying our individuality. We deserve to live our lives without constantly sensing that someone is judging us. Numerous privacy scholars have argued that privacy protects us from chilling effects. When we are alone or among people who respect our individuality, for instance, our friends or family, we hopefully can express ourselves without fearing disapproval (Solove, 2008; Rössler, 2005, p. 142-169).

Calo warns that the increasing deployment of robots that suggest they can judge people will increase the number of chilling effects we can expect to experience. A law enforcement robot, for instance, may coerce people to present themselves as well-to-do citizens to avoid attracting its attention. (Calo, 2014). Calo argues that less imposing robots will produce similar effects too. If a robot convincingly creates the impression that it prefers one thing over another, it may discourage its users from expressing conflicting opinions. Suppose a robotic companion, like the one I described in section 1.1., implied that people should enter romantic relationships with members of the opposite gender (Monae, 2022). This machine could discourage its users from discussing LGBTQIA+-related subjects during interactions to avoid displeasing it.

Calo speculates that we would feel constantly observed and judged if robots that simulate “the presence of another” became pervasive fixtures of our daily lives, creating ongoing chilling effects that threaten to “dampen creativity, skew our thoughts and actions toward the mainstream, and hinder self-development in much the same way as actual ubiquitous surveillance.” (Calo, 2009). As I stated earlier, Calo emphasizes that we cannot solve this problem by appealing to dominant conceptualizations of privacy. Although we may worry that someone will misuse our information after they have observed us through judgmental

eyes, they have already harmed us by inhibiting our behavior (Calo, 2009; Solove, 2008). Calo claims that this applies to robots too. Even if a robot that created a chilling effect did not transmit information to other parties, it would still deter users from expressing themselves. Privacy should protect us from intrusions of this kind. Thus, when someone (or something) chills our behavior without a valid justification, we suffer a privacy harm of a non-informational nature.

I will argue throughout this thesis that robots can create other privacy concerns aside from chilling effects that do not involve the misuse of information. Indeed, the literature reviewed in this section shows that robots can distract or discourage people from taking care of meaningful tasks by mimicking human social behavior, thus suggesting that we value being free from such disruptions. Aside from Calo, these contributions do not employ insights drawn from privacy scholarship to assess the ethical issues they raise. They imply that machines that simulate the presence of another human can disrupt people's day-to-day lives without recognizing that this may impact affected individuals' privacy. In the next section of this introduction, I will introduce theory that explains why we need to be left alone for the sake of our privacy.

1.3. Privacy as social disengagement

As I clarified earlier, contemporary accounts of privacy, especially those used to evaluate information communication technologies, tend to center around the use or misuse of information. When we consult the philosophical literature on privacy, however, it becomes clear that many privacy scholars would not accept this conceptualization as definitive. Furthermore, doing so would exclude crucial aspects of privacy from our definition of it. Most importantly for this thesis, we would fail to recognize that privacy also signifies our unwillingness to interact with certain people at times when we wish to focus on other activities. This section outlines theory from privacy scholarship that supports this claim.

Privacy scholars have maintained since the late 19th century that we must sporadically set aside our public commitments to live good lives. Coordinating with other people for prolonged periods takes a toll on us. Many, if not most, of the day-to-day social engagements we complete require us to conform to norms, standards, and expectations beyond our control. We often cannot attend to tasks that we find meaningful during these interactions. Being alone or among certain people of our choosing enables us to pursue needs and wants that we cannot

adequately realize in public settings. We need privacy, as to say, time dedicated to social disengagement, to enjoy aspects of our lives that we deserve to enjoy.

Samuel Warren and Louis Brandies were arguably the first scholars to recognize the value of being left alone. In their groundbreaking 1890 article, “The Right to Privacy”, the authors argue that people cannot endure “the intensity and complexity” of their public lives without periodically “having some retreat from the world”. Writing at the turn of the century in the USA, Warren and Brandies feared that industrialization, urbanization, and new means of surveillance would overwhelm people if left unchecked (Glancy, 1979). They contended that these ongoing socio-technological changes threatened to alter how American citizens related to one another for the worse. Cramped living conditions coupled with long work hours meant that many people of Warren and Brandies’ day were rarely, if ever, truly alone. Additionally, Warren and Brandies believed that bad actors, especially tabloid journalists, could now more effectively rob people of their privacy thanks to the invention of affordable, hand-held cameras (Glancy, 1979). The authors aimed to rectify this situation by advocating for a constitutional right to privacy.

According to Warren and Brandies, we can define privacy as “the right to be let alone”. They explain that we experience psychological distress when people bother or harass us after we have withdrawn from society writ large. If someone were to dog our every move, they would cause us serious emotional harm. Our private relationships, habits, interests, and communications would become public matters. We would quickly become exhausted by this. We would always sense that someone was nearby, therefore; dissuading us from completing our private affairs - many of which are crucial for our well-being. Warren and Brandies argued that the American judicial system should protect individuals from injuries of this kind by recognizing privacy as a legal right comparable to the right to liberty or property.

Over the past century, many other privacy scholars have contended that humans need relief from the social pressures they encounter throughout their day-to-day lives. Alan Westin, for instance, claims that every civilization that has ever existed has recognized that its members should distance themselves from others at certain times (Westin, 1964, p.24-30). He posits that modern liberal societies accept that people “must have substantial areas of interest apart from political participation” (Westin, 1964, p.35.). Our public commitments should not consume all our waking hours. Liberal theory stipulates that we deserve what we generally

call “free time” - during which we can enjoy “non-political” pursuits such as our hobbies and passion projects (Westin, 1964, p.35).

Furthermore, Westin argues that privacy affords people “emotional release”. He argues that sociological evidence shows that “life in society generates [tension]” (Westin, 1964, p.40) that culminates over time. We play various social roles that require us to temper our emotions and follow context-specific behavioral patterns when coordinating with others in public. For instance, we may switch from presenting ourselves as confident public speakers to deferential employees in a matter of minutes during work meetings. Moving from one role to the next drains us psychologically. We cannot fully express ourselves while playing social roles. And yo-yo between different psychological states when we interact with different people in different contexts. We cannot do this indefinitely. We must occasionally retreat from “the whirlpool of active life” to let off steam and regain depleted energy (Westin, 1964, p.41). We spend time alone or socialize with people such as our friends and family in private settings to recover from the “pressure of playing social roles” (Westin, 1964, p.41). Westin considers these periods of relief as essential for our well-being as they provide us with “a chance to lay [our] masks aside” (Westin, 1964, p.41) and freely express our emotions without fearing that others will judge us for doing so.

Julia Inness contends that privacy safeguards our right to choose with whom we share intimate relationships (Inness, 1994). She argues that many of the major decisions that shape our lives, for instance, whether we will get married or have children, alter the web of emotional ties we share with others. We assume new emotional responsibilities when we choose to become someone’s spouse or a parent. These decisions, Inness argues, should be private. The state, our employers, and people we do not love or like should not interfere with such choices because we cannot make them autonomously while being subject to public scrutiny (Inness, 1994, p.74-95). An example will help illustrate this. Until quite recently, same-gender romantic relationships were illegal in Europe and the United States of America. This private choice (e.g., with whom one wishes to spend their life) was a publicly decided matter. Thus, many people who would have entered same-gender relationships were prevented from doing so. Thankfully now, same-gender couples enjoy equal or similar rights to heterosexual ones ensuring that queer people can more freely choose their life partners.

According to Inness, our intimate and public-facing relationships are qualitatively distinct. Being a spouse, parent, or friend is quite different from being a competent employee or respectable citizen. The former set of relationships center around care, love or liking whereas the latter require us to set aside our personal needs to coordinate with usually non-intimate others as members of a group (Inness, 1994, p.95-116). Being among people who know us intimately enables us to express our vulnerabilities, feelings, and opinions more freely than we can in public - because these people care about such things and want us to be happy. If we could not spend time with people who love or like us, we would suffer grave harm as our ability to connect emotionally with others of our choosing would be severely limited, robbing us of something that we need to live good lives: intimacy.

Beate Rössler posits that spending time alone or with people we trust ensures that we can properly enjoy our autonomy. Although we can and do make autonomous decisions in public, distancing ourselves from others enables us to think more clearly about how we want to live our lives and what type of person we want to be. Devising life plans takes time, concentration, and energy, that we often cannot spare while focusing on tasks associated with our public commitments, for instance, our jobs. Furthermore, being among non-intimate others deters us from embracing our individuality and developing ideas that may seem heterodox or eccentric (Rössler, 2004, p. 43-71). Privacy shields us from these distractions, enabling us to plan our lives as we see fit.

We often consult people we trust when we wish to make decisions. Ideally, our friends and family want us to share our thoughts and feelings with them. We can communicate honestly and openly with these people as our closeness to them afford this. Hence, spending private moments with people we love or like lets us test out ideas that we would not feel comfortable discussing in public (Rössler, p.131, 2004). Suppose I was considering quitting academia because I believed pursuing another career path would make me happier. If this were the case, I probably would not tell my colleagues or superiors about my potential plans because I may strain my relationship with them by doing so. Instead, I would deliberate with my loved ones because I know they do not care about how I make money so long as I am happy. We need these types of private relationships to “invent ourselves” and flourish as individuals (Rössler, p.131, 2004).

Axel Honneth argues that our private relationships ensures that we develop the self-esteem necessary to realize that public actors should treat us with respect (Honneth, 2004, p. 18-25). When our friends and family members demonstrate through their words and actions that they value who we are, they show us that our subjective wants and needs are worth having. A good friend, for instance, will support us through thick and thin. If we realize that our jobs make us miserable because our colleagues and managers do not acknowledge our hard work, a friend will ideally tell us that these people have disrespected us. They may even recommend that we ask for a pay rise or look for a new position. Being told such things helps us recognize that we deserve what we want from life (i.e., decent working conditions) (Honneth, 2004, p.77-78, p.96-107).

According to Honneth, having supportive friends and family enables us to advocate for ourselves in public (Honneth, 2004, p.133). Our interactions with these people, which center around mutual love and care, teach us that individuals aside from ourselves can and should appreciate our subjective experience of the world. These relationships provide us with the means to recognize that non-intimate others should acknowledge that our needs and wants are valid. To return to the example I provided earlier, the friend in question would affirm our belief that our colleagues and managers should treat us better, thus affording us the validation and confidence we need to address this issue. We need support of this kind to deal with the challenges we face when we coordinate with people in public. Knowing that the people closest to us value our individuality - and believe that others should too – gives us the self-esteem required to pursue public-facing goals and prosper as members of society writ large.

Privacy thus allows us to pursue goals we cannot adequately realize when among people with whom we share non-intimate relationships (e.g., our colleagues or strangers). We need to spend time alone or with people we love or like to ensure that the manifold social interactions we complete throughout our day do not exhaust us and, ultimately, we can flourish as individual members of society. We value this aspect of privacy because it enables us to autonomously organize our lives and maintain mutually beneficial relationships of our choosing. If we were to frame this crucial aspect of privacy solely in terms of distribution of, or control over, information, we would fail to recognize why we value it. Having adequate control over one's information does not necessarily indicate that one "has privacy". For instance, we would be mistaken to say that someone who knows that no one can misuse their information yet lives their entire lives among people who could not care less about them can

enjoy their privacy (Marion Young, 1990, p.123-171). This person would be unable to withdraw from this dire situation and relate to themselves and others as one can in private.

When someone stops us from focusing on pursuits that require us to set aside our public commitments at moments when we should be able to do so, they deprive us of something we deserve, namely, privacy. Cataloguing the many ways this can happen is well beyond the scope of this introduction. Furthermore, much of this thesis aims to identify such issues and show how robots can produce them. I will argue that we should treat robots that distract, deter, or otherwise prevent their users from realizing their privacy, as discussed in this section, as causing harm comparable to a human who does the same. To reiterate the goal that I stated at the beginning of this introduction, I hope to steer the ethical discourse surrounding robots and privacy in a new direction through my work. And help other researchers, especially robot ethicists, see that privacy scholarship contains many valuable insights that enable us to identify privacy concerns that we would overlook if we uncritically subscribed to the dominant conceptualization of this value.

1.4. Chapter overview

I have divided this thesis into two sections. The first, which consists of chapters 1, 2 and 3, identifies under-researched privacy problems raised by contemporary robots. These chapters introduce theory rarely discussed within robot ethics - chiefly scholarship that treats privacy as safeguarding our ability to disengage from social interactions - to highlight issues we would fail to recognize if we relied on the dominant conceptualization of privacy outlined in this introduction. Although I do not usually state it outright within these chapters, I wrote them partly to push back against the ethical discourse surrounding robots that evaluates their effects on their users' privacy by examining whether they mediate information misuse. All three chapters present at least one privacy problem robots, that seem to understand human social behavior, produce that do not involve the transmission of information. I should emphasize too, that I constructed novel theory to discuss human-robot interactions and their impact on privacy to arrive at the ethical conclusions presented within these chapters. As such, I believe that they include an almost equal measure of theoretical and practical findings (i.e., ideas concerning what privacy is and how it is lost or preserved, respectively).

Whereas the first three chapters deal directly with privacy issues raised by robots, the final two are critiques that aim to improve how we, as scholars, discuss concepts that I implicitly

and explicitly refer to throughout this thesis. Chapter 4, for instance, reflects upon the normative significance of our homes and how domestic innovations (e.g., household robots) can change what we do inside them, potentially, for the worse – ideas I regularly discuss in the first section of the thesis without spending much time integrating them. And Chapter 5 uses the substantial sociological research I conducted throughout my PhD project to challenge an assumption I regularly encountered in the robot ethics literature, namely that humans and robots that observe norms produce ethically sound social interactions. I could not have written these last two chapters without developing the insights found in the earlier ones. However, one may read them in isolation as they indirectly deal with topics found in the earlier section of this thesis (e.g., why we must withdraw from others and the costs of social engagements, respectively).

In Chapter 1, I argue that social robots installed inside homes produce a novel privacy problem when they invite their users to engage with them. I synthesize two bodies of literature to accomplish this goal, namely: Erving Goffman's early sociological work on character performances, and privacy scholarship which argues that humans must sporadically withdraw from their public commitments to live well. Goffman and the privacy scholars I cite contend that humans spend much of their everyday lives attempting to meet socially determined expectations they do not directly control. Goffman calls such strategies "character performances" and clarifies that being among others who expect us to behave in such ways compels us to do so (Goffman, 1959, p.13-15). Furthermore, he argues that we may conceptualize the spaces we enter throughout our day-to-day lives as either "onstage" or "offstage" locales (Goffman, 1959, p. 109-141). The former represents places where we anticipate we must remain in character, while the latter symbolizes environments where we cease such effort to attend to tasks that we cannot adequately complete during performances. I draw from various privacy scholars' work to frame our homes as "offstage" environments which should enable us to attend to needs we cannot satisfy in public, onstage settings, which I label as self-care, autonomy, and intimacy. I argue that robots which elicit social responses from their users prompt them to perform in character as they would during interactions with the people they encounter in public, onstage setting, thus robbing them of time they could dedicate to the needs mentioned above – a phenomena I categorize as a privacy problem. The chapter introduces several core themes and ideas that I return to consistently throughout the thesis; specifically, the idea that robots can draw people into interactions, that our homes (and other private spaces) should shield us from unwelcome engagements, and that we may use

certain, foundational sociological constructs to discuss users' responses to a robot mimicry of human communication.

In Chapter 2, I again use sociological theory and findings to identify privacy problems robots produce when they simulate what it is like to interact with another human. In this case, a person one considers their friend. I begin by reconstructing an argument I call "the performance account" that other philosophers developed to advocate for the possibility of human-robot friendships. The performance account contends that Erving Goffman, and the sociologists he influenced, would support the idea that we classify someone as our friend when they consistently perform actions that communicate that this is the case, regardless of how they genuinely feel. Therefore, we can call a mindless, uncaring robot our friend if it behaves as we expect a friend to behave. I argue that this hypothesis does not hold water when we consult the sociological literature the performance account uses and that it excuses two privacy problems that convincingly friend-like robots will create. I interpret the sociological literature on "passing" to evidence these claims. I explain that people who practice passing conceal biographical facts concerning their marginalized identities during specific social interactions. Crucially for my argument, they do not always do this and usually tell people they wish to become friends with biographical information they conceal while passing. I introduce a thought experiment involving passing to demonstrate that we must tell someone something true about biographies to become their friend. As robots cannot experience anything, they cannot complete this action; thus, cannot be our friends. I conclude the chapter by demonstrating that friendly seeming robots will mediate surveillance by encouraging users to disclose sensitive biographical information during interactions - as they would when entering a friendship with a human and rob users of the care and compassion they deserve when they share truths about themselves that they only want their friends to know. I claim that both issues are privacy problems.

Whereas I focused almost exclusively on interpersonal aspects of privacy in the previous chapters, in Chapter 3 I examine the labor needed to produce spaces one may call "private" and robots' impact on these practices. I argue that we should expect robots designed to lessen the amount of housework homemakers complete to change how they perform such labor rather than reduce the time they spend on it. I appeal to numerous feminist scholars' work to demonstrate that we cannot enjoy the privacy afforded by our homes unless someone has produced the conditions necessary to do so, usually via unpaid housework. Women have

historically fulfilled this role and labored inside homes without financial compensation, thus leaving them with far less time to dedicate to their privacy than men. Over the past two centuries, many companies have created machines and devices that supposedly eliminate or reduce the need for certain types of housework. Feminist historians posit that these labor-saving technologies' widespread deployment has, ironically, increased rather than decreased homemakers' workloads. I identify three processes that lead to this outcome by interpreting several feminist historians' work, most notably Ruth Schwartz-Cowan's *More Work for Mother* (Schwartz-Cowan, 1983). I contend that such technologies: 1. compel homemakers to learn new, often highly sophisticated, and time-consuming skills to use them as intended; 2. create additional household responsibilities someone must assume; and 3. lead households and society writ large to expect more from their homes. I use these historical insights to evaluate a selection of contemporary robots, including Roomba, childcare robots, and Jibo. I argue that these machines will contribute to the processes outlined above, thus creating new housework, and changing what their users must do before they can enjoy their private lives. In the chapter's conclusion, I recommend that researchers abandon the widely held assumption that such robots save labor to ensure that their work accounts for the time and effort homemakers exert when using these machines to complete housework.

In Chapter 4, I and my co-author, Madelaine Ley¹, reflect upon a major normative assumption I leave mostly unquestioned throughout this thesis. Namely, that we deserve private spaces, most notably, our homes. We introduce three accounts of the home by interpreting philosophical and political literature dealing with this topic and show how ethicists of technology may use them to evaluate domestic innovations' moral import, including robots. We begin by discussing undoubtedly the more prevalent conceptualization of the home derived from classical liberal theory, we label "the relief account", which treats the home as a sanctuary cut off from the rest of the world, thus enabling occupants to recover the energy they spend coordinating with non-intimate others in public. Afterwards, we outline another well-known conceptualization of the home, we label "the production account", based upon the work of several Marxist-feminist scholars. This account, in contrast to the first, claims we should treat our homes as economic entities comparable to offices or factories as they provide

¹ I conducted the research and wrote most of this chapter. Madelaine Ley mainly helped to develop its core argument and with the writing of the section on Iris Marion Young's work.

occupants and, by extension, capitalist economies, with necessary resources including food, shelter, and childcare. Far from being cut off from the outside world, our homes are a necessary component of the political system we live under and should be treated as such. We assess both accounts' validity and highlight their deficiencies. We claim that the first fails to recognize that our homes can only function as sanctuaries because someone, usually women, has labored to keep them this way. And argue that the second reduces the care and intimacy we give and receive inside our homes into a commodity like any other. We develop a third conceptualization, inspired by Axel Honneth and Iris Marion Young's work, we label the maintenance account, to address the other two's flaws. We argue that our home lives should help us maintain self-esteem and therefore know that the people and institutions we encounter throughout day-to-day lives should treat us with respect. We identify two crucial factors that contribute to this outcome, namely, knowing that we share our homes exclusively with people who love us unconditionally (e.g., our friends or family) and have a space for ourselves that reflects our identity and histories. We conclude the chapter by showing how other ethicists of technology may use the maintenance account to evaluate domestic innovations.

In the final chapter of the thesis, I and my co-author Steffen Steinert² critique an assumption many robot ethicists and researchers from human-robot interaction studies have expressed. Namely, that humans and robots which observe norms produce ethically sound social interactions. We developed this critique because many authors from the fields mentioned above have called for the development of norm-compliant robots without properly examining how this could go wrong. We begin our critique by providing a sociological definition of norms that clarifies that this construct represents 1. largely unstated proscriptive and prescriptive rules that govern appropriate behavior during social interactions, 2. are socially determined and bound to specific groups, and 3. usually learnt pre-reflectively. We then interpret relevant sociological and political literature to outline what we call "seven troubles with norms" which we argue will derail efforts to build ethically sound norm compliant robots. We label these seven troubles, norm biases, paternalism, tyranny of the majority, pluralistic ignorance, norm abandonment and robot induced norm change. Overall, we claim that developing robots which observe norms will reinforce or reproduce social hardships and

² I wrote most this chapter. Steffen Steinert mainly helped to develop its core argument and with the writing of the sections "what are norms?", "outdated norms", "robot-induced norm change" and "conclusion".

ethical issues that many people already face. We conclude by recommending that researchers who wish to develop norm-compliant robots accept that doing so does not necessitate that these machines will seamlessly integrate into our social world and that we should develop means to ensure they do not make everyday interactions any more morally complicated than they already are. I chose to place this chapter at the end of the thesis as it contains many insights drawn from sociology and human-robot interactions studies I gathered during my PhD research which I could not present elsewhere, yet enabled me and Steffen Steinert to develop, what we consider to be, this much needed critique of the state-of-the-art.

2 Called Back Onstage: Dramaturgic Analysis, Domestic Social Robots, and Privacy

2.1. Introduction

So young Rossum said to himself: “A [human] is something that feels happy, plays the piano, likes going for long walks, and in fact, wants to do a whole lot of things that are really unnecessary.

Karel Čapek, R.U.R.

Erving Goffman claims that the vocabulary of dramaturgy includes many concepts that can be used to describe how humans coordinate during their day-to-day lives (Goffman, 1959, p.13-28). When we interact with other people in public settings, he explains, we follow behavioral patterns analogous to the scripts actors memorize to help them embody a character. While we have some room to improvise, we tend to pre-reflectively adapt our behavior to the expectations of others. We play numerous characters throughout our day, shifting from one role to the next as we interact with different people in different contexts. Goffman asserts that we cannot stay in character indefinitely. Actors take breaks during productions, and step offstage to attend to tasks they cannot satisfy while performing for their audience (Goffman, 1959, p.114-115). Likewise, we must distance ourselves from others to complete tasks unassociated with the characters we play in public (Goffman, 1959, p.116-120). These moments offstage often consist of short breaks, but also include longer periods spent in dedicated private spaces—most notably our homes.

I will argue in this paper that social robots designed for home use (henceforth domestic social robots) disrupt their users’ ability to remain offstage. These machines simulate the experience of interacting with other people by harnessing our tendency to treat things that resemble social actors like social actors. I will posit that these robots call their users back onstage when they initiate (artificial) social interactions. Furthermore, I will present an original synthesis of several privacy scholars’ interpretations of why we need time to ourselves to show that being called back onstage by robots is a privacy problem. I will argue that our homes should be places where we can set aside the characters that we play in public to pursue private activities related to our well-being, autonomy, and the maintenance of our intimate relationships. And,

being called back onstage by robots robs users of time they could otherwise dedicate to these tasks.

Overall, I aim to show that domestic social robots raise a privacy problem that scholars have yet to identify, that occurs when they invite their users to engage with them, and we should be wary of introducing them into homes for this reason. Before moving forward, I would like to ask my readers to view this paper as a conceptual rather than empirical exploration of an under-researched privacy problem made possible by social robots. I will draw from empirical evidence, when possible, but mostly rely on my interpretation of relevant sociological, human-robot interaction, and ethical research to build my argument.

2.2. Dramaturgic analysis

According to Erving Goffman, our lives in public are analogous to stage productions (Goffman, 1959, p 26-27). When we participate in communal activities, we perform as “characters” who others expect to behave in specific ways. Much like how we anticipate that an actor playing Macbeth will emphasize the tragic nature of this character through their speech and movements, we assume that someone working as a waiter will speak with their customers, take orders, and bus tables in a manner befitting this role (Goffman, 1959, p.22). We step onto a metaphorical stage when we engage with others in public settings and perform for onlookers, who serve as our audience. Meeting our audiences’ expectations ensures that they will treat us as competent individuals who have the qualities necessary to complete whatever task we are undertaking.

Goffman developed this way of conceptualizing social relations, called Dramaturgic Analysis, partly to explain why we perform seemingly unnecessary actions when coordinating as groups (Goffman, 1959, p.13-15). A waiter does not need to smile at customers while moving dishes to-and-from tables. Nor does a stage actor need to cup their face in their hands when reacting to shocking fictional events. These actions communicate that the waiter and actor can live up to, or exceed, their respective audiences’ expectations. Goffman explains that coordinating with others does not just call for us to mechanically complete tasks, but also requires us to present ourselves as the type of people who can perform such activities (Goffman, 1969, p.24; Goffman, 1966, p.24). I will draw from my experience working as a teacher to unpack these ideas.

When I enter a classroom, I step into the character of “a teacher” and follow behavioral patterns my students expect from someone playing this role. On some level, I appeal to a script by consciously and unconsciously reenacting successful performances I have seen other teachers and people in similar positions give in the past. I was never formally taught to present myself as a teacher, but I know through watching others and learning from my own performances how to conduct myself when around students. Although teaching formally requires me to verbally relay relevant information to my students, I perform many additional actions to communicate that I am good at my job. I use intonation that signals I am confident and pause between sentences to show I am reflecting on my words. I occasionally smile while giving presentations to ensure I do not seem unfriendly but generally keep po-faced to convey my concentration. I present myself as knowledgeable and attentive through these subtle yet highly meaningful gestures that are not essential for the task at hand yet expected from a teacher.

Stage actors who flub their lines or break character risk facing backlash from their audience. Failing to meet people’s expectations in day-to-day life results in similar outcomes. If I were to stare at the floor while teaching, I would have trouble garnering respect from students. Even if I fulfilled the formal requirements of my job (e.g., verbally relaying information to students), I would still risk upsetting students if I acted in this way. Although we avoid giving performances that will displease our audience, our actions are not necessarily motivated by fear of social backlash. Being among people who expect us to present ourselves as a character is often enough to prompt us to present ourselves as said character (Goffman, 1959, p.81-82; Goffman, 1966, p. 35-38). When students ask me questions during classes, they invite me to respond to them as a teacher would, something I do when performing this role, without much reflection. My tendency to automatically perform as a teacher when around students is plainly obvious when I run into them outside the classroom. Speaking with a student on the street, for instance, makes me pay attention to formalities associated with the classroom that I would not normally follow in this context. Without realizing it, I adjust my speech patterns and choice of words to match this social engagement, becoming the teacher this student expects me to be. And the teacher I want them to treat me as. We frequently step into character pre-reflectively throughout our day. Talking with cashiers in stores encourages us to perform as “customers”. Likewise, when communicating with our bosses we take on the role of “employees”.

Performing as characters, for Goffman, does not mean we are acting in bad faith. Our performances communicate that we understand (or do not understand) others' expectations of us. Like any other form of communication, we can convey truth or falsehoods via our performances (Goffman, 1959, p. 244-245). For instance, I enjoy teaching and truthfully express my enjoyment of this line of work through my performances in the classroom. Regardless of how we feel about the characters we play, Goffman explains, we cannot sustain performances indefinitely and must sporadically distance ourselves from others to attend to other aspects of our lives (Goffman, 1959, p109-141). For stage actors to give a good performance, they need time offstage to prepare themselves before their curtain calls. Once onstage, they must refrain from behaving in ways that would break the fourth wall. Goffman claims that this division between being on and offstage applies to everyday life as well.

Playing characters, Goffman explains, takes a toll on us (Goffman, 1959, p.129-141). We eventually become exhausted from having to adjust our behavior to meet our audiences' expectations. I cannot rest, pursue my hobbies, or answer phone calls while performing as a teacher, as this would communicate to my students I am not focused on the task at hand. However, I can complete these activities in other, appropriate contexts without anyone batting an eyelid. To complete tasks unassociated with the characters that we play throughout our day, we must step offstage and signal to others that we are taking time for ourselves (Goffman, 1959, p.109-141).

Goffman notes that workplaces often feature spaces where people can step offstage. Until recently, workplaces often had dedicated smoking areas where staff could socialize as friends rather than co-workers (Goffman, 1966, p.39-40). Likewise, restaurants usually prepare food behind closed doors, allowing chefs and waiters to act without worrying that they may accidentally offend their guests (Goffman, 1959, p.117-120). Even in highly public spaces, such as train stations, people can signal to others they wish to be left alone by picking up a newspaper or book (Goffman, 1959, p. 38-42). These spaces, Goffman claims, are analogous to a theatre's "backstage"—a place where actors drop character and take care of their personal needs without disturbing or (being disturbed by) their audience. He also claims that people usually know they should not enter these spaces without an invitation, much like how audience members during stage productions usually refrain from barging into backstage areas (Goffman, 1966, p.40).

Distancing ourselves from others is the most effective way to communicate that we no longer want to participate in performances. Solitude affords us space to complete tasks others may find inappropriate in public settings. Goffman suggests that we cast aside our public personas at our front doors (Goffman, 1966, p.9). Our homes shield us from social engagements. Their walls ensure that others cannot invite us to perform in character alongside them. Additionally, people usually know that they should not disturb someone when inside their homes, as they have clearly communicated that they want time off by removing themselves from the stage of public life. Aside from suggesting that performing for others eventually exhausts us, Goffman does not provide much normative justification for why we should be left alone at certain times. However, privacy scholars have argued that we must step offstage to fulfil several, crucial needs that cannot be adequately met in public settings—which I will detail later in this paper.

2. 3. Being called back onstage by robots

Researchers have known, for over twenty years, that people tend to treat simulated social stimulus as though it were the real thing (Nass & Moon, 2000; Calo, 2009; Turkle, 2011, p. 23-83). A computer program which speaks with a male-coded voice may sound more convincing than one with female-coded voice to anyone who implicitly or explicitly subscribes to gender stereotypes which portray men as better suited for leadership positions than women (Nass & Moon, 2000). Similarly, when a machine seems friendly and cooperative, its users will likely behave politely towards it and try to avoid hurting its non-existent feelings (Nass & Moon, 2000; Turkle, 2011). Social robotics, as a discipline, aspires to harness this tendency to create robots that simulate the experience of interacting with genuine social actors (Duffy, 2003; Breazeal, 2003).

Rather than creating technologies that one could plausibly call sentient, social roboticists usually aim to design robots that trigger the tendency outlined above (Duffy, 2003; Breazeal, 2003). It does not matter if users believe a robot has comparable psychological capabilities to a human, so long as its mimicry of social behavior prompts them to respond in kind. Roboticists often accomplish this by creating robots that mimic contextually relative behavior associated with specific and reasonably well-defined social situations (Hayes-Roth, 2004). For instance, people already know what to expect from service workers, thus a robot designed to help customers in retail environments should act like someone fulfilling this role (Dobrovestnova, Hannibal, & Reinboth, 2022). Thanks to our life-experiences, we know what

being in this social situation entails. When a service worker smiles at us, we understand that this is an invitation to engage with them and request their assistance (Hochschild, 2012, p. 3-12). And a robot that mimics this behavior will ideally produce a similar response from its users (Licoppe & Rollet, 2020).

When we interact with a social robot, we should not feel like we are dealing with a technical artefact. Whereas we must employ specialized technical skills to operate tools, machines, and most electronic devices, we should be able to rely on our social know-how to use social robots. By behaving like a human plausibly would in the same situation, social robots should motivate people to respond to their computationally determined outputs as though they were socially meaningful. In turn, this should encourage users to interact with a social robot by manipulating signs, such as spoken language or bodily movements, that they assume a human interlocutor would understand (Calo, 2009). Ideally, users can exclusively rely on communicative skills, otherwise used to coordinate with other people, to complete tasks with social robots.

Very few researchers have used Dramaturgic Analysis to describe human-robot interactions to date³. I intend to begin filling this literature gap in this paper and believe that Goffman's work contains many helpful concepts that we can use to interpret what happens when robots successfully convince people to treat them as social actors. Indeed, one could, and perhaps, should say that social robots invite users to perform in character for them.

To return to the service robot example again, this type of robot prompts its users to perform as "customers" in response to its mimicry of another familiar character from day-to-day life: a "service worker". Interactions between these two characters tend to follow predictable patterns. Service workers convey to customers their willingness to help them by smiling, making eye-contact, and greeting them with open questions (e.g., "how are you today?" or "is there something I can help you with?"). Someone who recognizes these opening strategies will likely respond to them as we expect a customer would. For instance, they may accept the service worker's invitation by maintaining eye-contact, reciprocating their smile, then politely make their needs known. Or decline their invitation by making it clear they do not

³ Mark Coeckelbergh is a noticeable exception and uses "performance metaphors" to describe how humans relate to technology (Coeckelbergh, 2019).

want assistance, perhaps by shaking their heads or raising one of their hands to signal “no, thank you” (Licoppe & Rollet, 2020).

A service robot should get its users to behave in this way. By convincing users to perform as customers, the robot encourages them to rely on their experience of similar social situations when interacting with it. The same holds for other social robots. Companionship robots mimic how we expect dependents to behave to ensure users perform as “caregivers” (Shibata & Wada, 2011). Likewise, an effective teacher-robot would have to make its users behave as “students” for it to function in this capacity at all (Sharkey, 2016). I should reiterate that we tend to step into character automatically when we enter familiar social situations. Social robots, one could say, trigger this type of response by mimicking contextually relative social behavior that, when performed by humans, call us onstage.

Being called onstage in public spaces is something we take for granted. We cannot go about our day without engaging with others who expect us to perform in character, as coordinating with others at our jobs, inside stores, or on the street demands this from us. Thus, robots that initiate interactions in these locales, arguably, do not disrupt our ability to remain offstage any more than a human would if they did the same. In contrast, we tend to treat our homes as places where we can disengage from performances. If a social robot were to invite a user to perform in character at home, this machine would call them onstage in a locale where we do not expect or, often want, this to happen. The robot would change this offstage locale into an onstage one.

In recent years, numerous companies have created social robots for home use that complete tasks which require interactions between at least two people when done by humans. These robots include embodied virtual assistants that perform household chores while maintaining an air of friendliness (Guizzo, 2015). Robots that serve as stand-ins for authority figures, such as nannies, tutors, or care workers (Okita & Ng-Thow-Hing, 2015). And, robots designed for entertainment purposes, which primarily function as artificial friends for children and adults (Turkle, 2011). Like other social robots, these machines invite their users to treat them as familiar characters from day-to-day life, in turn encouraging users to step into character themselves. A domestic social robot that acts like a subordinate (e.g., a housekeeper or assistant) encourages its users to perform as its boss and follow behavioral patterns associated with this character. Likewise, a domestic social robot that seems authoritative, may convince

its users to respond to it as though they were dealing with someone acting in this way (Calo, 2009; Calo, 2014).

Of course, peoples' responses to these invitations to step into character at home will vary. Whereas one person may immediately begin performing as an authoritative character when a domestic social robot asks for a command, another may dismiss this invitation. Nonetheless, the invitation is there, signaling to users that there is someone inside their homes who wishes to call them onstage. When domestic social robots send these signals, they disrupt their users' ability to shield themselves from social engagements by entering their homes—a phenomenon that, I will argue in the next section of this paper, amounts to a privacy problem.

2.4. Privacy as being offstage

Privacy scholars have drawn from Goffman's work to justify why we need time to ourselves since the 1960s (Westin, 2015). Although performing in character constitutes a significant part of day-to-day life, a life lived wholly among our peers, superiors, and strangers would be disastrous. Without periods of relief, we could not adequately take care of our well-being, make autonomous decisions, or cultivate relationships with people we love or like. In Western Liberal Democracies, our homes should provide us with space to pursue these needs without unwelcome disruptions (Warren & Brandies, 1890; Solove, 2008, p.58-61). Thus, when domestic social robots call users onstage, they intrude on their privacy. Before moving forward, I should state that I will not discuss privacy problems related to data misuse in this section. Instead, I aim to show, via my interpretation of several privacy scholars' accounts of why we deserve to be let alone at home, that being encouraged to perform in character by robots is a privacy problem.

Perhaps the first privacy scholar to appeal to Goffman in their work, Alan Westin, argues that we need privacy to attend to psychological and physiological needs that we cannot satisfy while among other people. Following Goffman, he explains that playing different characters throughout our day eventually exhausts us. We experience "tension" from having to meet the expectations associated with the characters we play, which at some point, we must attend to by finding relief in solitude (Westin, 2015, 41). When onstage, we often cannot attend to tasks required for our well-being. For instance, we cannot sleep, groom ourselves, or rest. Additionally, Westin highlights that liberal theory recognizes that people deserve time-off from their (public) social commitments to enjoy themselves and focus on their passions. To

live “a good life” we need “time devoted to sports, arts, literature, and similar non-political pursuits” (Westin, 2015, p.35). For Westin, we find fulfilment in activities unassociated with our public lives and therefore deserve leisure time to pursue them.

Other scholars have argued that being alone affords us a crucial right protected by liberalism: autonomy (Solove, 2008, p.31-37; Mokrosinska, 2018). As Goffman explains, we adjust our behavior to meet communally determined standards when playing characters; thus, have less freedom to express ourselves as individuals. Beate Rössler contends that privacy safeguards an aspect of our autonomy that we cannot fully realize in public (Rössler, 2007, p.43-71). According to liberal theory, everyone deserves to govern themselves as they see fit (Mill, 1985). As our lives are our own, we have the right to choose how we will live them. In practice, however, we have commitments to others that limit this freedom. Doing our jobs, for instance, requires us to conform to rules that we do not determine for ourselves. If we wish to maintain a good relationship with our bosses, clients, and co-workers, we cannot do as we please and must act in accordance with their expectations of us. In contrast, we can act with less concern for others’ expectations when alone. Withdrawing from others affords us a degree of autonomy usually unavailable to us. Private moments, often spent inside our homes, allow us to reflect on who we are, who we wish to be and formulate life goals without worrying that our actions or thoughts will upset the people we coordinate with in public (Rössler, 2007, p.79-106).

Even though being in public and, by extension, performing in character, influences what we can and cannot do during a significant part of our daily lives, this does not mean we always hide our true selves from others. When we wish to form intimate relationships with other people, we share personal information with them that we usually keep undisclosed (Inness, 1992, p.56-74). We let our guards down to build and maintain bonds with our friends, romantic partners, and family members. As Rössler states:

In private relationships - to the extent that they are private - we act differently, present ourselves differently, rehearse ourselves in a way that differs from what occurs in relationships with whom we are not on special or close terms. In this respect, the private sphere constitutes nothing less than a symbolic space in which, in our dealings with persons of our own choosing we can invent ourselves or at least act without protection. (Rössler, 2007, p.131)

These types of relationships require us to set aside performances. Being someone's friend means sharing private information with them, often concerning our thoughts and identities, that we do not share with everyone we meet. When we grant people access to our private lives in this way, we (hopefully) gain a trusted confidante whom we can step and remain offstage with. Without these relationships, we would have less leeway to be anything other than the characters we embody in public settings.

Goffman makes similar claims concerning the need to step offstage for the sake of intimacy. A married couple hosting a dinner party, for instance, cannot express the full extent of their love for one another while among their guests. First and foremost, they present themselves as ideal hosts by following appropriate etiquette and attending to other diners' needs. Once alone, however, they can let their guards down and honestly discuss how the night's events made them feel, thus enriching their relationship (Goffman, 1959, p.84).

Attending to these needs requires our more-or-less undivided attention, hence why we often withdraw to private spaces, such as our homes, that shield us from unwelcome social engagements, when we wish to satisfy them. We cannot adequately care for ourselves, experience the level of autonomy outlined above, or bond with people we love or like when performing as characters. Furthermore, the time we can dedicate to these needs is limited. Most working-age people in the Global North spend at least eight hours a day, five days a week, at work, where they must coordinate with others often in character (Penz & Sauer, 2019).

When domestic social robots call users onstage at home, it follows, they rob them of time they could otherwise dedicate to valuable, private activities. Say, for instance, a robot virtual assistant, interrupts a couple discussing their plans for an upcoming holiday to request commands from them. This request invites the couple to perform as the robot's superiors at a moment when they wish to enjoy their downtime together, bond over a mutual exchange of personal information, and make plans for their future. Being called back onstage by the robot distracts them from these valuable tasks, leaving them with less time to enjoy planning their holiday together. Furthermore, this happens inside a private space where we expect to be let alone for the sake of our well-being, autonomy, and intimate relationships, thus calling into question a home's ability to shield its occupants from unwelcome social engagements.

One could argue that anyone who experiences this privacy problem does so by choice. People do not have to buy domestic social robots and can refrain from introducing technologies that call them onstage into their homes. Although this is true, being called onstage is not something we usually notice happening. Furthermore, researchers have yet to frame this phenomenon as a privacy problem, meaning that it is not an issue that consumers or roboticists likely consider when buying or designing robots, respectively⁴. Therefore, consumers may purchase robots that invite them to perform in character at home without realizing that this will rob them of the precious time they can dedicate to their private needs.

Additionally, as stated earlier, social roboticists often aspire to create robots that simulate the experience of interacting with an authentic social actor. Whereas other people usually understand when someone wishes to be let alone, as far as I know, no robot to date has been designed to recognize the signals we use to communicate that we want privacy (e.g., avoidance of eye contact). If roboticists do not recognize that people need to be let alone at home, they may inadvertently create robots that encourage users to engage with them at moments when they wish to attend to private tasks that require their full attention.

2.5. Conclusion

I argued in this paper that social robots invite users to perform in character, as described by Erving Goffman. When this happens inside people's homes, I posited, robots call their users back onstage resulting in them losing time they could otherwise spend on tasks that require privacy, namely those associated with their well-being, (increased) autonomy, and intimate relationships. I claimed that this issue (being called back onstage) is a privacy problem, especially since it occurs inside a space that we use to shield ourselves from social engagements: our homes.

As stated at the beginning of this paper, the argument provided throughout is chiefly conceptual in nature. I have drawn from empirical evidence as much as possible but primarily relied on my interpretation of relevant scholarly literature to frame human-robot interactions as performances. Then, argued that performing with other people (and robots) takes a toll on

⁴ Ryan Calo does highlight a comparable privacy problem but focuses on how social robots may chill people's behavior rather than encourage them to perform in character (Calo, 2009).

us. Although one could view my strategy as overly reliant on speculation, I believe it lays the ground for new avenues of research that human-robot interaction scholars should pursue.

First and foremost, I believe researchers should pay closer attention to the subtle psychological costs of interacting with social robots. Coordinating with other humans demands a lot from us, even when we enjoy doing so. Knowing when, where, how and with whom we should perform as a given character, for instance, takes skill, effort, and learning. As made clear throughout this paper, employing our social know-how to interact with other people and robots is psychologically costly. Indeed, some scholars have even framed our usage of learnt social behavior as “labor” partly for this reason (Bourdieu, 2010; Hochschild, 2012; Penz & Sauer, 2019) As far as I know, human-robot interaction scholars rarely discuss this issue in their work and often assume that people prefer dealing with robots that respond to social behavior than ones which do not, without recognizing the skill and effort required to do so. As such, I highly recommend human-robot interaction scholars acknowledge and study the psychological costs of coordinating with robots as though they were humans.

Although I focused on domestic social robots in this paper, the conclusions presented apply to social robots in other contexts. We usually expect to have less privacy at work, on the street, or in retail environments, than at home. Nonetheless, we do need time offstage in these places too. As stated earlier, workplaces often include spaces where people can drop character. Likewise, we often signal to others that we do not wish to interact with them in public spaces via various means (e.g., by holding a newspaper or cell phone in front of our faces). If robots that call users onstage appear in these settings, they will disrupt people’s ability to remain offstage when they wish or need to—albeit less severely than ones installed inside homes. Considering that many social robots designed for workplaces and public settings have reached the market in recent years, I highly recommend researchers begin questioning whether these machines’ presence curtails the amount of offstage space available in these environments.

3 Friends Are Not “Electric” (Characters): A Sociological Case Against Human-Robot Friendships

3. 1. Introduction

All the world is not, of course, a stage, but the crucial ways in which it isn't are not easy to specify.

Erving Goffman, *The Presentation of the Self in Everyday Life*

I present a sociologically informed case against the possibility of human-robot friendships. Considering that many other researchers have already shown that robots cannot be our friends via more conventional, philosophical means (Sparrow & Sparrow, 2006; Elders, 2015), let me explain why I chose this somewhat unusual method. In recent years, several of my peers working in robot ethics have used the sociology of everyday life – especially the work of Erving Goffman – to argue that we may accurately call a robot our friend if it convincingly behaves like one, even though it is an uncaring, unthinking machine (DeGraaf, 2016; Coeckelbergh, 2017; Danaher, 2019). They imply that the sociological scholarship I just mentioned suggests that being a friend amounts to giving a character performance. According to Goffman (and the researchers he influenced) we can describe many interactions this way. Indeed, we can compare our collective understanding of what we should and should not do during social engagements to the techniques different stage actors use to portray the same character.

We often conceal our opinions and emotions during such performances to ensure we do not upset our audience. The scholars working in robot ethics I mentioned in the previous paragraph claim that our friends also behave like this. They complete actions we expect from friends and often do not communicate what they genuinely think or feel to avoid spoiling a performance. These scholars contend that if we accept that a human friend can express falsehoods about their psychological states yet remain our friend, we may say that a mindless robot that “consistently and coherently” (Danaher, 2019) behaves like such a person is our friend, too. I strongly disagree and will disprove this conclusion by drawing from the sociology of everyday life to show that robots cannot fulfil a necessary condition of friendship, namely, the disclosure of biographical information. As far as I know, I am the first scholar

from robot ethics to describe this necessary condition of friendship and identified it via my close reading of the sociological texts referenced throughout this contribution.

I chiefly employ sociological research that discusses the information management strategies stigmatized individuals use to evade discrimination to build my case. Specifically, I delve into the literature on “passing” – a form of character performance that enables marginalized people to conceal their identities to receive better treatment during interactions. People who practice passing falsely present themselves as members of a dominant group. While doing so, they cannot openly communicate biographical facts associated with their stigmatized identities without potentially outing themselves. Crucially, for my argument, they tend to deliberately stop passing to enter friendships. When a gay man, who otherwise passes as straight, willingly tells someone about his experiences being gay we can safely assume he wishes to relate to them as a friend. He tells them something true about himself that he does not reveal to everyone else. I argue that this holds for all friendships. We disclose biographical information to other people to become their friends. We stop playing characters while doing so. As robots cannot experience anything - let alone events that would shape their biographies – they cannot share such information with their users. Although they may convincingly simulate what it is like to interact with a friend, they cannot be one.

After I finish constructing the argument outlined above, I identify two privacy issues anyone willing to defend the claim that we can enter friendships with robots must excuse. Firstly, robots almost always transmit information they receive from their users to their manufacturers. We cannot trust them to behave like our friends when we tell them sensitive facts about ourselves and claiming otherwise would legitimize surveillance mediated via these machines. Secondly, if a robot successfully convinces someone to interact with it, as friends do, it could manipulate them into divulging deeply personal information that only someone who genuinely cares about them should hear. I conclude by outlining how we could use the sociological insights I present in this contribution to accurately describe why someone may mistakenly believe a robot is their friend.

3. 2. The performance account

Over the past two decades, many companies have created robots designed for companionship. Robots of this type currently operate in various contexts, such as people’s homes, care facilities, and schools (Sharkey, 2016; Guizzo, 2016; van Wynsberghe & Li, 2019). Examples

here include: Jibo, an embodied, household virtual assistant that appears to learn about its users' personality over time (Guizzo, 2016); Pepper, a humanoid machine that responds to outward displays of emotions (Carros et al., 2022); and Professor Einstein, an animatronic replica of its namesake which can chitchat with students while teaching them science (Kolodny, 2017). Robots of this kind, in general, provide companionship by mimicking behaviors we associate with human friends. Their manufacturers design them to fulfil this role and use various technical means to achieve this end.

Since the mid-2000s, numerous leading robot ethicists have argued that robots which appear friendly are deceptive. Robots cannot provide “care, companionship, or affection” because they are machines (Sparrow & Sparrow, 2006). They lack the mental qualities necessary to relate to someone as a friend would. If someone thinks otherwise, they have developed a false belief (Scheutz, 2009). Although someone may claim that a robot enjoys their company, thus, is their friend, this cannot be true (Turkle, 2011, p.53-67). A robot cannot enjoy spending time with this person. Whereas a human friend can and hopefully would (Elder, 2015). Robot ethicists, including the ones I just cited, tend to view robots which encourage people to treat them like friends as potentially or outright harmful. For instance, if someone mistakenly believes a robot can provide care, they may overestimate its capabilities and let it handle responsibilities that only humans should perform (e.g., childcare tasks) (Sharkey & Sharkey, 2020). Furthermore, we rob people of the truth when we let them believe a robot is their friend and, therefore, take part in a lie. (Sparrow & Sparrow, 2006).

Some robot ethicists disagree with this well-established position and contend that we can call a robot our friend without expressing a false belief. As stated in this contribution's introduction, they evidence this claim by categorizing friendships as character performances like those described by Erving Goffman. I will henceforth call this argument “the performance account” for this reason. Before I begin outlining this account in earnest, I should state that I constructed it by synthesizing claims I found in various articles and book chapters published over the past decade that use similar strategies to defend the possibility of human-robot friendships. Thus, I cannot attribute it to a single author. Nor do I wish to direct my criticism towards any specific scholar. As the performance account appeals to Erving Goffman's sociological work on character performances, let us begin by unpacking what this means. In the mid-twentieth century, Goffman developed a novel way of describing human interactions that used the language of dramaturgy to label social phenomena. Most famously, he compared

the communication strategies we use to convince others that we understand the largely unspoken, context-specific rules that govern everyday social engagements to stage actors' character performances (Goffman, 1959, p.13-28). From my reading of his work, he chose to use this analogy for following reasons.

First and foremost, we often embody something akin to a fictional persona while completing tasks that call for face-to-face communication. For instance, people who work in the service industry usually appear friendly and eager to please (Dobrosovetsnova, A., Hannibal, G., & Reinboth, 2022). They seem to be such people but, truthfully, have fabricated a persona to ensure that their interactions with customers follow a relatively predictable rhythm. Secondly, different people playing the same character will behave comparably. Although every receptionist, server, and flight attendant has an individual style of customer service, they collectively behave like other people who also work in this capacity. We expect this from them. A server who politely takes our order at a restaurant will not surprise or disturb us, whereas one who snickers at our menu choices probably will. We experience something similar when we attend stage productions. Every one of the thousands of actors who have played Macbeth over the past four centuries has interpreted this character differently. However, they all gave comparable performances once on stage that someone familiar with this play would recognize as Macbeth-like. Finally, we do not play characters all the time (Goffman, 1959, p. 109-141). Many situations do not call for such performances. Service workers do not remain in character during breaks or after their shifts have ended, because they do not need to create the impression that they want to help customers when none are nearby (Goffman, 1959, p.166-203). Goffman compares such periods to the time actors spend off-stage (Goffman, 1959, p. 109-141). Someone who plays Macbeth does not do so indefinitely. Even during productions, they will drop character to complete tasks backstage. For instance, they may change their costumes, speak candidly with co-actors and production staff, or simply eat a snack. If they did these things while their audience could see them, they would break the fourth wall and potentially ruin a performance.

Notice that my discussion of Goffman's work implies that people may communicate falsehoods concerning their perceptions during character performances. We often temper our emotions and keep tight-lipped to please an audience. A competent service worker will smile at a customer even if they dread talking to them, signaling geniality at moments when they do not feel this way at all (Goffman, 1959, p. 81-82; Hochschild, 1979; Hochschild, 2012, p. 3-

24). We cannot tell whether someone who knows how to play a character well has honestly conveyed their perceptions or feigned a psychological state to make sure our interactions with them go smoothly (Goffman, 1959, p.28, p.203-231). Many service workers enjoy their jobs and sincerely express their emotions by cheerfully greeting customers. However, many others dislike what they do for a living and effectively lie when they create the impression that they want to help someone. We cannot see inside other people's minds; thus, we will treat a service worker who loves their profession and another who hates it similarly if they behave like one another. We encounter situations like this all the time (Goffman, 1959, p.28, p.203-231). During lectures, university students will react comparably to an instructor who genuinely feels confident and one who fakes this emotion to impress them (Coggins, 2023). Likewise, an office manager who praises a subordinate because they appreciate their hard work and another who does so because they know they should pretend to care about such things will generate the same response from said employee (Penz & Sauer, 2019)

The performance account interprets friendships along these lines. Our friends draw from a repertoire of techniques to maintain the impression that they like us (Coeckelbergh, 2010; Coeckelbergh, 2017; Danaher, 2019). They nod along attentively while we tell them about our day and laugh when we share humorous anecdotes with them. If we feel sad, they will comfort us and perhaps even offer us advice. They communicate that they care about us by performing actions that we collectively use to convey this message. They demonstrate that they understand what being a friend entails by behaving like other people we may call our friends (Coeckelbergh, 2010; Danaher, 2019). According to the performance account, they play a character (Coeckelbergh, 2010; Coeckelbergh, 2017, Danaher, 2019). As such, they sometimes mislead us to meet our expectations. Our friends often tell us white lies to spare our feelings or pretend to listen when we talk about something they find boring (DeGraaf, 2016; Coeckelbergh, 2019, p.126). We cannot know if our friends' words and actions genuinely reflect their intrapersonal experiences or amount to dishonest signals sent to preserve a performance (Danaher, 2019). Demanding complete honesty from our friends would make it impossible for them to play this character well (Coeckelbergh, 2010). We probably would not call someone who tells us everything that goes through their mind our friend because they may say things that we would prefer they kept to themselves. If someone we considered our friend desperately wanted to announce they hated how we dressed, we would expect them to suppress this urge. We know that our friends regularly deceive us. We have all committed such deeds at some point in time. Doing so does not void a friendship. On

the contrary, we usually act this way to ensure that whoever unknowingly witnesses our lies will continue to see us as their friend. The performance account contends that this observation proves we can truthfully call robots our friends.

Suppose a robot consistently performs actions like those described above (Danaher, 2019). Someone may understandably believe that this robot cares about them because it looks and sounds like a human who wishes to convey this message. We understand that such information does not always reflect our friends' perceptions. Nor can we ever know for sure that it does or does not. We can only relate to the outward displays of friendliness our friends impress upon us. We consider someone who competently cultivates this impression our friend regardless of their mental state (Coeckelbergh, 2010; Coeckelbergh, 2017, Danaher, 2019). Therefore, we may treat a robot that does the same as our friend without expressing a false belief. Claiming otherwise would disqualify our human friends from being our friends. Prohibiting them from miscommunicating their thoughts would fundamentally change the relationship we share. We expect them to lie occasionally to stay in our good graces (DeGraaf, 2016). If we accept that we remain friends with humans after they perpetrate such deceptions, we must concede that we can enter friendships with robots that convincingly communicate that they like us despite feeling nothing towards us. Hence concludes my reconstruction of the performance account.

3. 3. What is passing?

I contend that Goffman's work and the sociology of everyday life, in general, does not support the performance account's conclusion. According to this body of literature, being someone's friend does not amount to playing a character. Goffman implies this himself (Goffman, 1959, p. 109-141, p. 166-203; Goffman, 1963, p. 31-45; Goffman, 1966, p. 39-42). However, his work does not directly discuss this subject. As such, I will interpret additional scholarship, most of which Goffman influenced, to evidence this hypothesis. I construct my argument by examining a specific type of character performance called "passing". People who practice passing miscommunicate their identities. They anticipate that being honest may produce negative, potentially dangerous responses from others because they belong to a marginalized group. They obscure their identity behind a performance. Furthermore, they tend to deliberately disclose sensitive, biographical information they conceal while passing to make friends. Doing so ends their performance. I will return to this crucial point in the next section of this contribution after I have explained what passing is.

Let me introduce some presuppositions before I begin my discussion on passing. Firstly, I do not want to imply that people should or should not practice it. I solely aim to describe it. I consider it something some marginalized people do, usually because they understandably fear discrimination. I do not need or wish to examine the political or moral implications of passing to develop my argument. Nor will I. Secondly, I will frequently reference social categories that constitute people's identities in the following two sections. I will treat these things as social constructs. One becomes gay, straight, cisgender, transgender, disabled, or able-bodied by living within a society that differentiates people based on these historically and culturally relative distinctions (Foucault, 1975, p. 135-195; Stone, 1987; Oliver, 1990, p. 43-92; Butler, 2007, p. 107-175). If someone primarily expresses traits associated with being heterosexual, we will probably classify them as such. Likewise, if someone communicates that they only share romantic or sexual relationships with people of the same gender, we will likely conclude that they are gay. Such categories only make sense in societies that interpret sexuality this way (Akroyd, 2017, p. 1-7). Europeans have not always treated same-gender attraction as signifying one's "homosexuality". Indeed, we began grouping people according to their orientation relatively recently (Thorp, 1992). Every category we use to denote social difference has a history too. They do not represent eternal, essential qualities that some people have, and others do not. Instead, they reflect how our society currently orders us.

This final point helps us understand why people practice passing. Ordering often implies a hierarchy. Many individuals treat some groups worse than others (Hill Collins, 2000, p. 243-291; D'Ignazio & Klein, 2020, p. 24-49). Despite the substantial efforts to curb homophobia in Western Europe over the past fifty years, people still regularly experience bigotry during day-to-day interactions when they communicate that they are gay. Simply holding one's partners' hand on the street can be dangerous for gay people. Likewise, talking openly about one's orientation in public may anger homophobes within earshot. Other marginalized groups experience similar forms of oppression. Transgender people, as a demographic, suffer harassment and assault at an alarmingly high rate (Stryker, 2008, p.203, p.221-224; Faye, 2020, p.1-17). While disabled people often endure patronization, unwanted attention, and ridicule while going about their day (Siebers, 2004; Cox, 2013). Living under the threat of the types of interpersonal oppression outlined above is exhausting. Hence, marginalized people have developed many strategies to avoid entering hostile social situations. These strategies include passing.

Goffman discusses passing in his 1963 book *Stigma: Notes on the Management of Spoiled Identity* (Goffman, 1963, p. 85-92). He explains that we infer people's group affiliations by interpreting verbal and non-verbal symbolic representations of their biographies (Goffman, 1963, p.68-80). A man who speaks about his ex-boyfriend signals that he has had a romantic relationship with another man. Likewise, someone who has a visible prosthetic limb shows that they have acquired or were born with a disability (Goffman, 1963, p. 57-64). In both cases, these people reveal biographical facts about themselves that disqualify them from being considered heterosexual or able-bodied, respectively. We cannot rewrite our life histories. If we have done or experienced something that affiliates or disaffiliates us with a group, others will classify us accordingly once they learn about these events (Goffman, 1963, p.129-151). People often do not rank others based on such things. I hope my readers agree that queer, straight, able-bodied, and disabled people all deserve respect and fair treatment. Nonetheless, prejudices still run rampant in the modern world. A homophobic individual will react antagonistically upon discovering that a male interlocutor has an ex-boyfriend. Just as a person who thinks disabled people are weak or defective will assume that someone who uses a prosthesis has these character flaws. Prejudiced people make these judgements after they realize someone's biography proves they belong to a group they consider deviant, dangerous, or incapable (Goffman, 1963, p. 57-64). They cannot arrive at this conclusion without sufficient evidence.

Some marginalized people conceal such evidence to convince others they have biographies they do not have. They effectively create disguises by manipulating symbols associated with groups that do not face discrimination (Granfield, 1991; Kanuha, 1999). Goffman and other sociologists call this practice passing. Sometimes passing requires people to communicate falsehoods through spoken language. For instance, a gay man may pass as straight by fabricating stories about female love interests when someone asks him about his dating history (Brown, 2010; Renfrow, 2011). Whereas at other times, it demands physical effort from them. For instance, someone who uses a prosthesis may obscure it underneath their clothing to pass as able-bodied (Goffman, 1963, p. 92-125). Some things are easier to hide than others. We usually assume that people who communicate that they exclusively find members of the opposite gender attractive are straight. Thus, we probably will not doubt a gay person who makes such claims. In contrast, we cannot always control how others interpret our outward appearances without refraining from entering some social situations altogether (Garfinkel, 1967, p. 116-186). A person who uses an artificial limb cannot disguise this fact during

activities where they need to disrobe or wear clothing that does not cover their entire body (e.g., sexual intercourse, medical examinations, or sporting events). There are countless ways people may accomplish passing. However, practicing it always requires them to suppress or downplay traits that announce their marginalized status by looking and sounding like someone who does not have these traits (Allport, 1955, p. 145-146, p. 150-152; Goffman, 1963, p.85-92; Kanuha, 1999; Renfrow, 2011).

Regardless of how or why someone passes, they must vigilantly manage the information they impress upon others. One slip of the tongue or misstep may ruin an otherwise convincing attempt to pass. A seemingly heterosexual person who mentions they frequent a gay bar may raise eyebrows while speaking with people familiar with this establishment. Similarly, we will realize someone we assumed was able-bodied has a physiological disability if they neglect to hide their medical aids during an interaction. People cannot pass all the time. They cannot persuade someone who knows about the facts they obscure while passing to interpret their assumed identity as credible. Once they disclose information of this kind to someone, they stop passing (Garfinkel, 1967, p. 116-186; Stone, 1987; Rogers, 1992). This sometimes happens at inopportune moments. Being outed as gay during interactions with evidently homophobic people would endanger someone. At other times, though, passing would complicate or disrupt an interaction that does not call for it. A disabled person probably will not present themselves as able-bodied during appointments with medical professionals who help them manage their condition. Indeed, doing so would prevent them from talking openly about their health and medical history with this person who already knows they have a disability (Siebers, 2004). There are many other situations that motivate people who practice passing to communicate honestly about their experiences, including the interactions they share with people they trust will respect their identities. I will return to this point in the next section of this contribution.

Although Goffman does not state it outright, his description of passing resembles his account of character performances. Considering that he more-or-less stopped using the language of dramaturgy to describe social interactions by the time he published *Stigma: Notes on the Management of Spoiled Identity*, this is somewhat unsurprising. Nonetheless, if we used this framework to discuss passing, we could say that someone who practices it plays a character. I am not the first scholar to suggest this. Indeed, numerous sociologists have appealed to Goffman's work to make similar claims (Kanuha, 1999; Renfrow, 2011). Passing requires

people to communicate information via their words, appearances, and behavior to conform to others' expectations. When done well, it enables practitioners to enter and complete interactions without encountering unwelcome disruptions. Additionally, people who pass often miscommunicate their perceptions. They behave like unhappy servers who begrudgingly smile at guests while taking their orders. Or our friends, when they pretend to care about our troubles despite having something else on their minds, according to the performance account. They express falsehoods about their experiences because they know that honesty would ruin their performance and potentially expose them to discrimination.

3. 4. Friends are not “electric” (characters)

In the previous two sections of this contribution, I briefly discussed something that the performance account overlooks. According to Goffman and other sociologists who draw from his work, we do not play characters during every interaction. Using this framework to interpret all the social situation that we encounter throughout our day-to-day lives would misrepresent what it is like to be among other people. Goffman mentions many interactions of this kind in his dramaturgically focused work but does not spend much time examining them (Goffman, 1959, p. 231-249). Nonetheless, he heavily implies that the moments we share with our loved ones, including our friends, do not center around character performances (Goffman, 1959, p. 84; Goffman, 1963, p. 155-164; Goffman, 1966, p.3-13). I do not expect my readers to treat my interpretation of Goffman’s work as gospel truth. As such, I will use the sociological insights I introduced earlier to prove my point.

I build my case by presenting a thought experiment involving passing. The narrative I construct below illustrates that we must stop attempting to manage another person’s perceptions of us through a character performance to become their friend. I chose to explain this process via a discussion on passing because people who practice it indisputably drop character when they deliberately disclose biographical information concerning their marginalized identities to bond with someone they trust. They cannot play the character they were playing once they have communicated such information. I argue that we always behave this way to initiate friendships. We must tell someone something true about our experiences to let them know us as friends know each other. We do not play characters while doing so. Although I describe the experience of making friends, I will not offer a comprehensive definition of being one. Instead, I aim to identify a necessary condition of friendship that robots cannot satisfy, Namely, the truthful communication of biographical information.

Suppose a gay man chooses to pass as straight at his office job. He has come out to his friends and family. However, he does not want his colleagues, managers, or clients to learn anything about his private life that could out him. For the sake of argument, let us say he experienced discrimination at a previous workplace. He conceals his orientation because he does not trust the people who he encounters during work hours to respect his identity. He refrains from discussing life experiences at work that someone could interpret as symbolizing his attraction to men and changes the subject whenever someone asks him about his love life. He has one colleague, though, he likes more than others. He enjoys working with this person and often spends his lunch break chatting with them. Additionally, this colleague has consistently shown that they respect gay people and suggested that they understand the risks of being out at work. One day he decides to tell this colleague that he is gay.

Why would someone share such deeply private information with someone else? As the saying goes, one cannot unring a bell. This action will forever change how these two people relate to each other. We can safely assume that the protagonist of this thought experiment believes he can trust his colleague. He senses that telling them about his identity will not upset, enrage, or alarm them. Instead, he anticipates they will appreciate this gesture and understand they should not disseminate the information they learned. Let us say they do. They pretend they never heard this secret and behave as though they believe their colleague is straight to ensure he can continue passing. They know something that other people at their workplace do not. If they revealed this information at work, they would simultaneously out their colleague and massively damage - or utterly ruin – their relationship. I think my readers will agree that these two characters have become friends; or, at the very least, gone through an event together that may produce this outcome. Indeed, I contend that I just described a necessary condition of friendship.

Something important happens when we willingly disclose biographical information in the manner sketched above (Rössler 2007, p.129-141). We invite someone to become our friend. We let them know us differently than they did before by telling them truths about ourselves that we do not disclose during many other day-to-day interactions (Inness 1992, p.95-116; Rössler 2007, p.131). If they recognize and welcome this invitation, we may earn a friend. This does not always happen. Sometimes we tell people things they do not want to hear which make them think worse of us. For instance, the fictional gay man introduced in this section

could have mistakenly believed his colleague did not harbor homophobic views. If this were the case, he would have received a markedly unfriendly response to his invitation. Thankfully, his colleague's perception of him did not deteriorate after he came out. If anything, it improved. They recognized that he had entrusted them with information they must keep to themselves and supported his decision to pass. This does not automatically mean they will become friends though. Once they have spent more time together, these two characters may discover that they do not have much in common or disagree upon matters they both find important. Afterwards, they may decide to part ways and end their budding friendship. Nonetheless, they laid the grounds necessary to become friends. One of them shared biographical information, while the other respected what it meant and retained it. We cannot make friends without completing the first action in this sequence.

Let me reiterate that I chose this "high stakes" example for illustrative purposes. We can substitute the information revealed in this thought experiment - I would argue - with any biographical fact that someone does not disclose indiscriminately. For instance, one person may fear that others will judge them if they learn about their struggles with addiction, whereas another may worry that their peers will find them crass once they discover they enjoy an unpopular hobby. The content of this information matters less than people's expectations of its owner. We know that we cannot tell every person we meet everything about us without turning otherwise unremarkable interactions into fiascoes or, quite possibly, disasters. A server who recounts traumatic childhood memories while taking orders will probably annoy their guests and may even face the sack if they turn this indiscretion into a habit. Likewise, a gay man who passes as straight at work to avoid discrimination may incense or infuriate his bigoted colleagues if he discusses his dating history. We strategically choose with whom we share such information. Under the right conditions, this action will lead to the formation of a friendship (Inness, 1992, p. 95-116).

I will now interpret the thought experiment sketched above via social theory that portrays interactions as character performances. Let me begin with a question. When the gay man who practices passing at work came out to his colleague, was he playing a character? I would answer, no. He was playing a character beforehand. He managed his colleagues' perception of him to ensure their interactions remained agreeable. Maintaining this performance while coming out would be impossible. He cannot appear straight to his colleague after he says he is gay. He did something utterly out of character. He may soon recommence his efforts to

present himself as straight during interactions with his colleague, especially when someone from their office is nearby. Nonetheless, he had to stop doing so to communicate true biographical information. At least for a moment, he let his colleague see behind the performance he uses to protect himself. He dropped character to attempt to bond with them. I contend that we must always do this to enter friendships. We do not play characters when we tell someone something true about ourselves to invite them into a friendship. We reveal an aspect of our biography to them that they may like or dislike. Either way, this action signifies the absence of a character performance.

Please allow me to speak directly to my readers for a moment. I assume you have made a new friend in the past, perhaps quite recently. You may have met this person at work, at a party, or online. During your immediate interactions with them, you may have "played a character" as described throughout this contribution. Spending time with new people, especially ones who we want to like us, can be daunting. Hence, we often rely on character performances to influence others' perceptions of us. You may have tried to convince this potential friend that you are the kind of person with whom they would enjoy being friends by exaggerating or concealing traits you have. You may have appeared more fun, sensitive, energetic, or sociable than you actually are. Perhaps you laughed at a joke they made that you did not find particularly funny or expressed gratitude when they gave you a gift you did not want at all.

At some point, though, I am sure you behaved like the protagonist of the thought experiment I introduced in this section. You communicated something to this person that accurately reflected what it is like to be you. Maybe you told them about your strained relationship with your parents. Or mentioned that you wish you could quit your job to pursue your real passion. You stopped trying to impress this person, as one does during a character performance, to let them know you as friends know one another. This is how we initiate friendships. Robots cannot do this. They cannot know, understand, perceive, or feel anything. They lack experiences altogether. Certainly, they could convince someone otherwise by saying or doing things that suggest they have lived through events that shaped who they are. Nonetheless, this will always amount to fiction. They cannot invite someone into a friendship by telling them something true about themselves because they do not possess such information. Therefore, they cannot be our friends.

3. 5. Privacy issues

I developed this argument against the possibility of human-robot friendships because some philosophers have drawn an antithetical conclusion by appealing to the same body of literature I referenced throughout this contribution. I labelled such efforts to prove that we can call robots our friends without expressing a false belief “the performance account”. The performance account claims that being a friend amounts to playing a character, thus we can become friends with robots that behave this way. In contrast, I have shown that we cannot enter friendships with someone (or something) that creates the impression that they are our friends until they stop playing a character, at least momentarily, to share biographical information with us. Even the most convincingly friend-like robot (e.g., one that behaves precisely as we expect friends to behave) cannot fulfil this condition. As friendly as it might seem, such a machine cannot disclose information that accurately reflects its experiences – because it has none.

Although I chiefly aimed to critique the performance account via a close reading of the sociology of everyday life, I have frequently alluded to something that ethicists would call a moral value. Namely: privacy. I implied that friendships require privacy. We tell our friends facts about ourselves that we do not tell everyone. We let them know us as the people we are rather than the characters we play (Rössler 2007, p.129-141). It would be well beyond the scope of this contribution to explain exactly why we value the privacy we share with our friends. Nonetheless, I will conclude my critique of the performance account by identifying two privacy issues it excuses by claiming that we can relate to robots as though they were human friends.

Firstly, if someone genuinely believed a robot was their friend, they would almost certainly share biographical information with this machine to try to bond with it. This is how we build friendships with humans after all. Thus, we can assume that people who want to become or remain friends with a robot will behave comparably. Although I have almost exclusively focused on the first time this happens in this contribution (i.e., the moment someone initiates a friendship via information disclosure), we continuously tell our friends things about ourselves that we do not disclose to everyone (Inness, 1992, p. 74-116; Rössler 2007, p.129-141). The content of such information ranges widely. As suggested throughout the previous section, however, we tend to share biographical facts with our friends we suspect many other

people would find displeasing. Obvious examples here include: our sexual orientation, mental health histories, or dislike of our jobs. Whereas we expect people we already or wish to consider our friends to understand they should not disseminate such information, we cannot trust robots to do this.

Many, if not most robots, continuously transmit information to their manufacturers. For instance, robots that mimic human speech constantly upload audial recordings to their manufacturer's servers for processing (Sharkey & Sharkey, 2010; Kudina, 2021). Someone who shares biographical information with a robot, therefore, may unknowingly share this information with a company too. Many end-user agreements stipulate that companies can use data of this kind as they wish (Terpstra et al., 2019) and research shows that people usually do not read these contracts before they begin using technologies that gather data from them (Solove, 2012). Hence, someone who believes they have a robot friend may inadvertently disclose deeply personal information to said robot's manufacturer that they only want their friends to know. Furthermore, technology companies regularly sell datasets concerning their user bases' preferences, identities, and habits to other businesses and, sometimes, governmental agencies (O'Neil, 2017; Zubuff, 2019). Supposed robot friends, it follows, almost certainly would mediate surveillance in some capacity. The performance account overlooks this privacy issue. Claiming that robots that outwardly appear friendly are friends, even though they cannot stop sharing the information they receive, would simultaneously gloss over these machines' surveillance capabilities; and suggest that our human friends may share the often highly sensitive information we tell them without jeopardizing our relationship.

A proponent of the performance could counter this argument by claiming that we should only relate to robots as friends when they do not transmit such information through the internet. A well-meaning company, for instance, could create a robot friend that preserves its users' privacy by ensuring it does not send them any information deemed private. Its users may interact with this robot without worrying that unseen parties will learn anything revealing about them. I posit this would still create a privacy issue – albeit one that requires several argumentative steps to identify.

Many people would react negatively to the facts we disclose to our friends. For instance, our bosses do not want to hear about how boring we find our jobs and we may upset a stranger if

we spoke candidly about our mental health while waiting with them at a train station. Sharing information of this kind is risky and we can create unpleasant or hostile situations when we mistakenly relate to someone as a friend. We become vulnerable when we tell someone our true thoughts, feelings, and experiences because they may learn things about us that they do not like. If someone were to behave like a friend and successfully encourage us to tell them such sensitive information, even though they did not care about us, we would probably feel like they wronged us somehow. Suppose they create the impression that they want to hear about our experiences dealing with an aspect of our lives that causes us stress so they can help and support us as a friend would. If they told us after this exchange that they do not like us nor care about our struggles, we could claim that they manipulated us into divulging information that we only want our genuine friends to hear. Furthermore, they convinced us to let our guard down and take a risk that we would not have taken if we knew their true nature. Even if this pretend friend did not disclose what they heard to anyone else, they nonetheless deceived us to gain access to information only someone who legitimately cares about us should know.

I contend that the hypothetical, privacy-preserving robot I outlined earlier would produce a harm comparable to the one created by this pretend friend. A robot that encourages its user, via its friendly demeanor, to share information in the manner sketched above prompts this person to take the leap of faith required to initiate a friendship. A robot cannot understand the significance of this action. Indeed, only other humans can. Thus, this person acts courageously for a machine that does not care about them or recognize the vulnerable state they have placed themselves in. A human friend would (ideally) treat this action with the respect it deserves and likely feel closer to this person afterwards. In contrast, a robot “friend” would make this person feel like they have bonded with someone when they have not, thus, depriving them of the care, compassion, and affection they deserve when they expose themselves to potential social backlash for the sake of a friendship.

Although this may not sound like a privacy issue, largely because it does not involve the wrongful transmission of information, I contend that it is. Indeed, the process described above conflicts with a principle derived from the legal and ethical literature on privacy, namely: *we deserve to share our private lives exclusively with people of our choosing who love or like us* (Warren and Brandies, 1890; Inness, 1992, p.106; Solove, 2008, p.34-35; European Court of Human Rights). Such people include our friends. The performance account fails to recognize

this. Indeed, it suggests that robots or people who appear friendly are our friends, thus, do nothing wrong when they encourage us to disclose our vulnerabilities to them under false pretenses. When we tell someone (or something) information that we do not typically disclose to other people, we communicate that we have chosen to form a private relationship with them. If they create the false impression that they want the same thing, they disrespect us as individuals who deserve to enjoy our private lives with people who care about us.

3.6. Conclusion

I began this contribution with two objectives in mind. Firstly, I aimed to demonstrate that the sociology of everyday life does not support the idea that we can become friends with robots - despite what some robot ethics have claimed. And secondly, I wished to identify two privacy issues that robots which convince people to treat them like their friends would produce. Furthermore, I believe that I have introduced sociological insights that lay the grounds for new avenues of research while working towards these goals. Indeed, I think I can provide a preliminary answer to an important question I have yet to address based on the content of this contribution. Specifically: why might someone believe they have a robot friend according to the sociology of everyday life? Let us return to the discussion of service work I used to introduce Goffman's social theory in section 2 to think this through.

As I stated, service workers often create the impression that they like or care about their customers. This is a crucial aspect of their profession (Hochschild, 1979; Penz & Sauer, 2019). Many of them take vocational courses to learn how to maintain an air of friendliness even during highly stressful situations at their workplace. An expert service worker will manipulate symbols (e.g., their tone of voice, facial expressions, and choice of words) to convince people that they enjoy interacting with them regardless of their actual mood. I assume that my readers know this. We generally understand that service workers who appear friendly are just doing their jobs. Sometimes, though, people do not recognize this fact. They become convinced that a server, receptionist, or flight attendant wants to share a private rather than professional relationship with them (Urry, 2005, p. 59-74). They mistake an impression of friendliness for the real deal.

We could say that something similar happens when someone develops the false belief that they have a robot friend. In both cases, a person erroneously thinks that someone or something wishes to bond with them because they have misread symbols that would communicate this

message in other settings. They failed to notice context cues that would have prevented them from making this mistake (e.g., the fact they are interacting with someone who is paid to help them or talking to a machine). I believe this comparison accurately captures how a sociologist of everyday life would interpret this phenomenon. Nonetheless, developing the conceptual and theoretical means to adequately use this observation to describe why someone could confuse a lifeless, uncaring robot for a friend would merit another contribution (at least). I hope to return to this topic soon. Until then, I would like to invite other researchers from robot ethics to develop the ideas presented above themselves because I believe that sociological research of this kind would enrich philosophical discussions on human-robot interactions that involve the simulation of friendliness.

4 More Work for Roomba? Domestic Robots, Housework, and the Production of Privacy

4.1. Introduction

They something something all their lives. Work like robots. Yes, that would fit. They work like robots all their lives.

Ira Levin, *The Stepford Wives*

In their 2020 work, *Data Feminism*, Catherine D’Ignazio and Lauren F. Klein call upon their readers to ‘make labor visible’. They explain that the technology industry all too often fails to give credit where credit is due. Every line of code, electronic device, and statistic we use today exists thanks to the labor of dozens, if not hundreds of people whose work tends to go unrecognized. Take, for instance, Amazon’s Echo, one of the world’s most popular smart home devices. While Amazon, the company, claims ownership of this product, it would never have reached the market without the work of innumerable people. To produce a single Echo unit, Amazon sources material and services from myriad global supply chains, relies on in-house engineers to create code and employs factory workers to assemble the device into a finished product (Crawford and Joler, 2018). Each step in this manufacturing process requires Amazon to capitalize on people’s labor. Yet, whereas Amazon’s employees hopefully receive fair compensation for their work, many of the other people the company depends on do not, including the miners who extract the minerals needed to create the Echo’s electronic components (Crawford, 2020) and the ‘ghost workers’ the company contracts to help train its algorithms (Gray & Suri, 2019).

D’Ignazio and Klein claim that we should strive to make labor visible to ensure we can identify whether the technological resources we rely on were produced under fair working conditions and judge these facts accordingly. In recent years, researchers from robot ethics have begun to apply this type of thinking to their objects of study. Whereas proponents of robotization tend to frame this process as diminishing the need for human labor (Schwab, 2017), numerous scholars have highlighted that introducing robots to pre-existing work environments changes how workers perform labor instead of simply decreasing their workloads.

Nurses working alongside healthcare robots, for instance, must learn new skills to ensure these technologies function correctly and do not jeopardize their ability to care for their patients (Van Wynsberghe & Li, 2019). Likewise, service robots create new standards of affective labor by projecting an air of perpetual friendliness, which service workers must emulate to satisfy their customers (Dobrosovetsnova, Hannibal, & Reinboth, 2021). And rather than helping construction workers to do their jobs, installing robots on building sites may result in their human operators spending more time cleaning up after these machines than focusing on tasks they know how to complete themselves and find professionally meaningful (Muishout et al., 2020). Failing to acknowledge the additional labor these people perform to accommodate robots does them a disservice. Moreover, doing so would leave an important question unanswered. Specifically: is it even worthwhile for these workers to adapt their professional practices around robots?

This article will shed light on the labor required to produce something, via robots, that is not generally considered a commodity or service; the conditions necessary to enjoy our private lives at home. Over the past two decades, numerous companies have created robots that promise to provide their owners with more free time by automating domestic tasks that require skill and effort when performed by humans. Having robots help out around the house may seem like an attractive option to anyone who has a busy schedule, as these machines appear to decrease the amount of work required to keep homes in agreeable states. In this article, however, it is argued that delegating domestic tasks to robots alters the nature of housework rather than reducing the need for it. While they may streamline some tasks, employing robots to create an environment at home conducive to private activities, such as bonding with loved ones, rest, or leisure pursuits, takes work. To use these robots as intended, users must learn new skills and take on new responsibilities. Additionally, it is argued that users may excuse a robot's failure to complete household tasks because they enjoy interacting with them.

Historical accounts of older domestic technologies, which will be drawn upon throughout this article, demonstrate that the processes outlined above are nothing new. Indeed, some scholars have convincingly argued that domestic technologies designed to reduce labor result in their users performing more housework overall while making it appear as though they are doing

less (Strasser, 1982; Schwartz Cowan, 1985). Domestic robots may represent the continuation of this trend.

The article proceeds as follows. In the second section, it defines what is meant by domestic robots and discusses how their supposed ability to ease labor inside the home helps explain their popularity. These machines offer their owners something that has been promised in liberal thought arguably for centuries: a private life free from labor. In the third section, a critique is provided of older domestic technologies (based largely on Ruth Schwartz Cowan's work) that have prevented these types of technologies from delivering on their labor-saving promises. Historical insights are developed to examine several domestic robots and illustrate how they reshape housework. The conclusion is that producing a home ready for private pursuits has always required labor and that domestic robots will not remedy this state-of-affairs.

4.2. Housework and the production of privacy

Recent estimates suggest that there are over 30 million domestic robots deployed worldwide (International Federation of Robotics, 2019). Although they perform various functions and possess varying degrees of sophistication, these robots share several characteristics. First, unlike robots designed for industrial purposes, domestic robots do not create commodities with exchange value, but provide their owners with something more akin to a personal service. Anyone who owns a Roomba, for instance, cannot sell what this robot produces – because it does not produce anything that can be sold. Instead, they rely on it to relieve them from a household chore they would prefer not to do themselves, namely, vacuuming (Fortunati, 2018). Secondly, domestic robots operate inside people's homes. Many domestic robots were designed to do so. In contrast, others were initially manufactured to work in other settings (e.g., education, healthcare, or the service industry) but have since been introduced to the home. And thirdly, they are commercially available, meaning that individuals or families can purchase them from retailers if they have enough money.

Domestic robots offer their owners something that, ideally, should already be provided to them in modern, liberal democracies: a private life free from labor. Since the late nineteenth century, liberal theorists have framed the home as a sanctuary where individuals can withdraw from the pressures of their work lives and govern themselves as they see fit (Gobetti, 1997; Rössler, 2007, pp.23–7). Working for a living requires us to conform to the impersonal norms

of the marketplace (Prost, 1998). Although we can hopefully negotiate the terms of our employment, we must fulfil commitments to others to exchange our labor for an income. Maintaining a good relationship with our employers or clients demands that we rein in our individuality. If we wish to continue working with these people, we cannot do as we please. Instead, we must deliver goods or services on time, follow codes of conduct, and behave professionally. We cannot express the full extent of our individuality at work, as our occupational commitments restrict what we can and cannot do.

Our lives would be intolerable if we were always subject to workplace rules and norms. We need time to ourselves to attend to private interests which we cannot fulfil at work. Under liberalism, we have the right to pursue activities essential for our well-being and happiness once we finish our workday. These private activities include maintaining close relationships with people of our choosing, recreational pursuits, such as hobbies or sports, and self-care (Westin, 1968; Inness, 1992). Being guaranteed time off from work ensures that we are not overwhelmed by our jobs and can enjoy aspects of our lives unassociated with our labor relations. In liberal thought, our homes serve as the polar opposite of our workplaces and represent the most private locations available to us. Once we go through our front doors, we can expect a level of privacy unobtainable elsewhere (Solove, 2008, pp.58–61). What we do at home does not concern our employers or anyone else we coordinate with at work. Indeed, we enjoy special legal protections inside our homes that allow us to go about our private business without being intruded upon by others (Warren & Brandeis, 1890; Council of Europe, 1950).

Despite being among the most fundamental tenets of modern liberal thought, in practice, the ability to keep our private lives free from labor amounts to an ideal rather than a guaranteed right. Realizing the value liberalism attaches to the home requires a substantial amount of work in itself. Since the 1970s, feminist scholars and economists have criticized the idea that labor stops at our front doors. The home, these scholars argue, produces many resources needed to keep the wheels of the market turning. In 1977, for instance, Scott Burns estimated that households generate approximately one-third of United States gross national product as they collectively supply the workforce with essential resources, including shelter, food, education, childcare and community services (Burns, 1977; Kumar, 1997). Though necessary for any modern economy, these resources are generally produced without financial

compensation. Cooking for one's family is not something that we can exchange for wages. Nor will entertaining restless children at home secure us an income.

According to such feminist scholars as Silvia Federici and Catherine MacKinnon, these unpaid domestic activities (among many others) revitalize the workforce (MacKinnon, 1991; Federici, 2012). If no one attended to their upkeep, our homes would soon become disagreeable places where we would struggle to enjoy our private lives and likely have trouble returning to the demanding world of work. Dinner does not spontaneously appear on the table, nor do floors miraculously clean themselves. Someone must take care of these labor-intensive domestic tasks to ensure household members who have paid jobs can return home to regain their strength after clocking off for the day. Historically, women have supplied most of the labor required to maintain households and usually undertook this work without payment. As such, the home has always been a workplace for many people, albeit one without fixed working hours, holidays or pay.

Gender continues to play a significant role in influencing whether someone will labor at home for free; however, it is necessary to note that men and people of other genders face similar disadvantages to women when working as homemakers. Every moment we spend on housework, regardless of our gender, means we have less time for private activities we find valuable or essential for our well-being. Even if households divided housework among themselves fairly, these tasks still need to be completed. For many, if not most, people, housework is a reality of modern life and something that needs to be done to enjoy the benefits of their homes.

Having robots take care of housework for us seems like a suitable response to the issues outlined above. Delegating household tasks that we do not wish to complete ourselves to machines would provide us with more time to focus on activities we find valuable. With their help, we could produce the conditions required to enjoy our private lives at home with far greater ease. This line of reasoning helps explain why domestic robots have become so popular over the past two decades. However, domestic robots alter, rather than diminish, the labor needed to realize the value that liberalism attaches to the home.

4.3. A brief critique of domestic technologies

Far from being the first technologies meant to save labor inside the home, domestic robots are but the latest additions to the long list of products manufactured for this purpose. The modern home includes many devices and machines designed to make homemaking more manageable. For example, vacuum cleaners allow their users to clean floors more efficiently than they could with a dust-pan and brush. Likewise, refrigerators ensure their owners can keep their homes stocked with food that would quickly spoil if stored in a cupboard or pantry, thus reducing the number of trips they make to grocery shops. Even though technological innovation has eliminated the need for some forms of housework, this does not mean that maintaining a home today requires less labor than it did in the past – as I will demonstrate in this section.

Despite usually being considered consumer goods, domestic technologies are not end-products that households purchase for consumption. Instead, they share more similarities with intermediate goods, such as industrial equipment or office computers, which facilitate the production of other resources (Kumar, 1997). For example, people do not usually buy vacuum cleaners because they find these machines aesthetically pleasing or intend to use them for recreational purposes. Instead, they invest in vacuum cleaners to produce a clean home. Many other widely used domestic technologies fulfil similar functions. A refrigerator has little worth beyond keeping food ready for someone to turn into meals. And washing machines let their owners clean their clothes without taking them to laundrettes, where they would have to pay for this service. These technologies play a role within production processes, enhancing their users' ability to meet household needs.

Historians have questioned whether homemakers' workloads have decreased over the past two centuries as a result of the proliferation of mass-produced domestic technologies. Joan Vanek, for instance, calculated that the number of hours American women spent on housework remained relatively stable from 1924 to 1974 (Vanek, 1974). Vanek and other scholars have hypothesized that domestic technologies developed during this period (many of which are still in use today) provided homemakers with the means to produce more for their families rather than saving labour time (Vanek, 1974; Bittman et al., 2004). For example, a homemaker with a washing machine installed inside her home could keep her family supplied with freshly laundered clothes throughout the week. Producing this outcome, however, meant

that homemakers did the laundry more often than before. Whereas they once dedicated a slot in their workweek to manually scrub, soak and wring clothes, after the washing machine's introduction, they started doing laundry whenever possible, thus:

[Changing] the laundry pile from a weekly nightmare to an unending task, increasing the size of the pile, the amount of water and fuel and laundry products most households used, and possibly even the housewife's working time, which was now spread out over the week. (Strasser, 1982, p.588)

According to Ruth Schwartz Cowan, adapting pre-existing practices to accommodate domestic technologies tends to neutralize their labor-saving effects. History shows, she claims, that technological leaps forward inside the home reconfigure housework, resulting in homemakers taking on new tasks in response. Throughout her seminal work on industrialization's impact on American domestic life from the nineteenth century onwards, *More Work for Mother*, Schwartz Cowan argues that female homemakers' duties consistently grew during this period. The increased availability of affordable, mass-produced domestic technologies contributed significantly to this outcome (Schwartz Cowan, 1985).

Schwartz Cowan details two centuries of industrial and domestic history to support this conclusion, which, for brevity's sake, will not be recounted here. Instead, the main mechanisms that led to homemakers, with access to ostensibly labor-saving domestic technologies, paradoxically undertaking more work will be established by focusing on one critical example, the cast-iron stove. Schwartz Cowan suggests the cast-iron stove is 'the single most important domestic symbol of the nineteenth century' and continues to serve as the primary source of heat in many kitchens throughout the world today (Schwartz Cowan, 1985, p.54). After the cast-iron stove became commonplace in the mid-nineteenth century, average American households changed how they prepared meals in three significant ways⁵.

⁵ These three sub-sections, detailing Cowan's account of the cast-iron stove, draw from chapter 3 of *More Work for Mother: The Ironies of Household Technology from the Open Hearth to the Microwave* (Schwartz Cowan, 1985).

Changing work processes

In *More Work for Mother*, Schwartz Cowan explains that housework usually involves multiple steps that people complete to produce a desired outcome. She calls these sequential actions “work processes”. For example, before households began relying on cast-iron stoves, the work process they followed to prepare hot meals centered around open hearths. An average dinner during this time consisted of meat, vegetables, grains, and water stewed in a single pot. Cooks, who were almost always women, prepared these ingredients, then left them to boil on their house’s hearth. Aside from stirring these stews occasionally to prevent them from burning, cooks could leave these one-pot meals unattended until they were ready to serve.

Cast-iron stoves changed all this. Thanks to their in-built ovens and numerous hobs, cast-iron stoves enabled cooks to prepare more complicated dishes than were possible with an open hearth. Although this meant that cooks could make more nutritious, varied meals for their families, to produce this outcome, they had to abandon an earlier, less laborious work process and master a completely new method of cooking. Preparing meals with cast-iron stoves requires that cooks understand how to control its various heat sources, know when to remove and place pots, and remain fixed behind this appliance for safety’s sake. This new work process became the standard way to make meals after cast-iron stoves overtook open hearths in popularity during the mid to late nineteenth century.

Changing responsibilities

Households have divided labor among their members, often based on gender, since humans transitioned from nomadic to sedentary lifestyles (Lerner, 1987; Engels, 2010). According to Schwartz Cowan, this holds for American working-class families in the nineteenth century. However, women’s and men’s responsibilities in the home changed significantly during this period. In the early part of the century, when households still used open hearths for cooking, men were responsible for gathering and preparing wood to burn on these fires and keep them stoked. Women specialized as cooks and stewed meals on the fires their husbands had made beforehand. Schwartz Cowan claims that men’s responsibilities for meals differed from women’s but were not necessarily easier. They often had to forage for and chop wood themselves, then spend time kindling and fanning flames to create a fire suitable for cooking.

After cast-iron stoves replaced open hearths as American working-class families' primary cooking apparatus, the division of labor outlined above disappeared. Because these stoves burned charcoal or gas, men no longer needed to collect fuel before their wives could cook, which meant they could withdraw from the kitchen altogether. In contrast, women took on new responsibilities. For one, they became the sole operators of their homes' primary cooking apparatus, as cast-iron stoves afforded a style of cooking that required someone constantly to monitor their various heat sources. Furthermore, they became far more adept cooks as cast-iron stoves allowed them to bake cakes, roast meats, and boil vegetables simultaneously – tasks which demand a considerable amount of skill and experience to perform.

Changing expectations

Schwartz Cowan suggests that domestic technologies, once widely deployed, can lead people to expect more from their homes. The cast-iron stove allowed cooks to make complex dishes that would have been incredibly difficult to produce on an open hearth. In the early nineteenth century, only wealthy families could regularly afford to eat meals more complicated than one-pot stews. Culinary goods, often made by skilled artisans, such as leavened cakes, unpreserved fresh meat, and multi-course meals, symbolized affluence. Buying and cooking these dishes or ingredients was beyond the means of most households as they lacked the money, equipment and skilled labor required. By the end of the nineteenth century, however, socioeconomic changes resulted in food that was once prohibitively expensive and difficult to produce at home becoming standard fare for working-class families.

As a result of industrialization, ingredients that were previously scarce and costly became affordable and widely available. Having access to these goods meant that households with cast-iron stoves could make meals that years earlier were consumed almost exclusively by the upper class. For instance, cakes containing eggs, white flour, sugar, and other flavorings became typical desserts that cooks often baked inside a cast-iron stove's oven while preparing savory items on its hobs. Because they now had the means to produce complicated food at home, people started to expect more from home cooking and acquired a taste for meals and dishes that took considerably more labor to make than one-pot stews. Of course, someone had to meet this new demand and almost invariably this task fell to women, who began spending more time in the kitchen to satisfy their families' new culinary expectations. Throughout *More Work for Mother*, Schwartz Cowan examines many more domestic technologies that altered

pre-existing work processes, created additional household responsibilities, and shifted people's expectations of their homes.

The changes to housework brought about by cast-iron stoves, and many other now ubiquitous domestic technologies, simultaneously made it possible for households to enjoy their private lives in new ways while producing, as Schwartz Cowan is wont to say, 'more work for mother'. The ability of working-class families to realize the value that liberalism (even at this time) attached to the home, improved after the Industrial Revolution, partly thanks to technical innovation. But these improvements came with a cost which female homemakers had to pay with their labor, thus ensuring that their lives at home bore little resemblance to the liberal conception of a private life.

4.4. Domestic robots and housework

Though much has changed since the conclusion of the Industrial Revolution, the need for house- work has remained a constant. Will domestic robots amend this state of affairs? Considering the historical precedent set by other domestic technologies, it seems unlikely. Indeed, we should expect domestic robots to reshape rather than diminish housework. We have been here before; domestic robots, like older domestic technologies, change how households go about realizing their private lives. It is beyond the scope of this article to provide an exhaustive review of the ways domestic robots change housework. Nonetheless, we can already see this process unfolding by analyzing the human–robot interaction.

This section begins with an examination of the changes to housework produced by the namesake of this article, Roomba. Whereas there is solid empirical evidence to show that Roomba does alter housework, the two other cases presented in this section are more conceptual in nature. As such, one should read the first case as an example of how domestic robots have already been shown to change housework and the second two as conceptual explorations of how this may happen.

Changing work processes

Roomba robotic vacuum cleaners reached the market in 2002 and are almost certainly the best- known domestic robots currently in production. Although it is difficult to ascertain how many people worldwide own one or more of these robots, Roomba's manufacturer, iRobot,

claimed in 2016 that 20% of the world's vacuum cleaners were robots and their own models accounted for 70% of the market (Etherington, 2016). As one would expect from a robot vacuum cleaner, Roomba can move around flat environments while sucking up dirt and dust without direct supervision. Users can command their Roomba to clean different areas of their homes via a smartphone app and schedule the robot to start vacuuming whenever suits them. iRobot markets Roomba as a product that takes care of its users' cleaning needs and allows them to 'forget about vacuuming for months at a time'⁶.

Roomba promises to streamline the process of cleaning one's home. It vacuums for its users, ensuring that they do not have to remember to monitor how dirty their floors are, schedule time for vacuuming, and perform this task themselves. The number of steps required to obtain the desired outcome of having dust and dirt-free floors seems to decrease thanks to Roomba. Although these robots do eliminate some aspects of this work process, users must perform new tasks to integrate Roomba into their cleaning routines and ensure it functions properly.

Roomba, it turns out, has difficulty traversing its primary place of deployment, the home. Because of its design and reliance on three wheels, Roomba cannot travel over many objects commonly found on floors. Electric cables and stray clothes are impassable obstacles for the robot because its chassis is too close to the ground to clear them. Additionally, Roomba cannot sense its size and often tries to squeeze through narrow or low spaces that cannot accommodate it. A regular dining room chair, for instance, will trap the robot between its legs, thus immobilizing it until someone comes to its rescue (Sung et al., 2007). These design flaws necessitate users modifying their homes to suit Roomba (Forlizzi, 2008)

Multiple ethnographic and human–robot interaction studies have shown that Roomba users develop new housekeeping practices to compensate for the robot's shortcomings. Using the robot as intended requires arranging homes in a Roomba-friendly manner by clearing away anything that might block its path. Adapting one's home to Roomba is not a one-time task; it is an ongoing process. Users must remain vigilant and remember that Roomba will fail to complete its cleaning schedule if they forget to tidy their homes appropriately beforehand (Forlizzi & DiSalvo, 2006). Even when users have taken these precautions, they frequently

⁶ Quoted on the iRobot website available at <https://www.irobot.co.uk/deals> (accessed August 2021).

have to rescue their Roomba when it gets stuck (underneath a chair or cabinet) (Sung et al., 2007).

Although users generally seem to appreciate having these robots vacuum their homes (Forlizzi & DiSalvo, 2006), it is uncertain whether they save any time at all. Indeed, Roomba cannot clean floors as effectively as a human with a manual vacuum cleaner (Vaussard et al., 2014). Its circular shape means it cannot vacuum corners where dust tends to settle, and its motor is not powerful enough to suck up heavier pieces of dirt. Even with Roomba's help, users still need to monitor their home's cleanliness and manually vacuum or sweep up dirt the robot has missed. Furthermore, users cannot rely on the robot to deep-clean particularly dirty parts of their homes and must take care of this task themselves (Sung et al., 2007). Does Roomba relieve its users from housework? On the one hand, users can clean their floors less frequently with Roomba. But on the other hand, they must remember to perform a whole series of new tasks and regularly clean their floors manually. Indeed, users often spend as much time cleaning their floors after purchasing a Roomba as they did before they owned one (Sung et al., 2007).

Changing responsibilities

Whereas Roomba's shortcoming prompts users to reconfigure pre-existing work processes, using robots for other tasks may lead to the emergence of household responsibilities that did not exist before their deployment. More specifically, delegating childcare tasks to robots may result in parents spending more time monitoring their children's ability to manage their emotions.

In modern, service-based economies, workers use emotional labor to earn a living (Penz & Sauer, 2019). Being good at our jobs often requires us to manage our feelings and learn to suppress or promote certain emotions to fit our employers', clients', or colleagues' expectations (Hochschild, 1979). A university lecturer cannot express frustration or boredom while giving a class. Lecturers must temper their emotions to convey that they find the task at hand engaging even when they do not. The same holds true for many other professions, especially those that involve face-to-face communication. Experienced waiters know how to respond to demanding customers. Likewise, people in leadership positions often undergo training in dealing with the emotional pressures of managerial work. Although we develop

skills of this kind throughout our lives, our emotional education generally starts at home with our parents (Hochschild, 2012). Parents teach their children norms and practices to prepare them for the outside world (Federici, 2012), including the emotional skills needed to coordinate with their peers, superiors and eventually, employers (Hochschild, 1979). Of course, different parents have different parenting styles, and people generally teach their children emotional skills that align with their own understanding of how one should feel, based on their experience and socialization (Hochschild, 2012; Bourdieu, 2010).

Over the past decade, numerous companies have created robots for home use that interact with children on a “social level” (Darling, 2012). These robots range from interactive toys which appear to develop personalities over time to anthropomorphized robotic playmates which provide children with companionship (Okita & Ng-Thow-Hing, 2015; Turkle, 2017). By manipulating signs, such as spoken language or body movements, these robots act as stand-ins for humans or pets and keep children occupied even when they are alone. Although parents might not see entertaining their children as housework, it is time-consuming and delegating tasks of this kind (to some extent) to robots may seem like an attractive choice for caregivers with other commitments (Sharkey & Sharkey, 2010). Letting children interact with these social domestic robots, however, is not cost-free and may result in parents taking on new responsibilities if they wish to prepare their children for life in public.

People can form emotional attachments to robots that appear life-like (Scheutz, 2009; Nyholm, 2020). Children, in particular, have trouble understanding that robots are disinterested machines and often treat them as though they deserve care, respect and affection (Turkle et al., 2006). Sherry Turkle, for instance, claims that children tend to believe that robots are alive or real enough to justify forming emotional relationships with them that they do not extend to other artefacts (Turkle, 2011). According to Turkle, children often do not recognize robotic toys as toys at all. Instead, they see robots as something akin to companions or dependents that have emotional needs. Turkle claims that children may come to prefer the fictional relationships they have with robots over those they share with humans, especially since these machines cannot get bored, frustrated, or distracted (Turkle, 2011). Unlike humans or pets, these robots cannot respond with displeasure or impatience when a child acts demandingly. Nor can they grow tired of conversations or play sessions. Thus, a child cannot learn important life lessons related to emotional management from these robots, such as when

they should graciously accept that an interlocutor has lost interest, or diplomatically suppress frustration during interactions that they find uninteresting.

There is very little research on how social robots affect children's emotional development. Nonetheless, indirect evidence does suggest that children can learn behaviors misaligned with their parents' wishes by interacting with social technologies. For instance, researchers working on the social effects of virtual assistants, such as Amazon's Alexa, have raised concerns that children may develop habits ill-suited for human-to-human communication by using these technologies at home (Wiederhold, 2018; Kudina, 2021). Users tend to speak with virtual assistants in short, direct commands, chiefly because these technologies rely on voice-recognition software that has trouble interpreting language aside from precise instructions (Ureta et al., 2020). Indeed, these technologies encourage their users to ignore valuable aspects of interpersonal communication, such as politeness, courtesy, and attentiveness (Burton & Gaskin, 2019). Whereas adults usually know that they should use this type of language only when speaking with virtual assistants, children do not; therefore, they may learn to undervalue language that signals they are listening and responding empathetically with their interlocutor. A child who speaks with virtual assistants, and by extension, interacts with social robots designed to entertain them, does not need to manage their emotions as they would if they were speaking with a human, as these technologies will always respond positively to them.

We can infer from the information outlined above that parents who wish to prepare their children for relationships outside the home and allow them to interact with social robots may have to keep a more watchful eye on their offspring's ability to differentiate between simulated and real social stimuli. Many parents would want their children to treat the relationships they project onto robots as equivalent (or preferable) to the ones they share with humans, as the former does not prepare them for the emotional management they must perform to coordinate with others outside the home. While letting children have make-believe relationships with robots may be harmless if parents make sure their children understand that this amounts to fantasy, this nonetheless entails the creation of a new household responsibility that did not exist before the introduction of these machines.

There are technical fixes available to address these problems. For instance, one could design robots for children that express emotions and can signal disapproval. Children can learn that

their feelings cannot always come first (Wiederhold, 2018). This, however, would provide robots (or more accurately their manufacturers) with the power to decide when children deserve to experience negative, emotionally loaded evaluations (Sharkey, 2016). Would this make things easier for parents? Parents would likely still need to monitor whether these machines acted in accordance with their parenting styles to ensure they do not negatively affect their children's emotional development.

Changing expectations

Households adapt existing practices or create new ones to accommodate domestic robots. Domestic robots may also shift peoples' expectations of their private lives. More specifically, households may come to value domestic robots for their perceived personalities and overlook the machines' shortcomings because people enjoy interacting with them. People frequently relate to robots as though they deserve to be treated like humans or animals. Humans, some researchers have argued, are hardwired to react to artefacts that resemble living beings as if they were the real thing (Calo, 2010; Turkle, 2011, p.8; Nyholm, 2020). Even robots that barely look or behave like people or animals can elicit responses of this kind. For instance, Sung and colleagues found that many of the Roomba users they interviewed named these robots and interpreted their algorithmically determined movements as signs of their personalities (Sung et al., 2007). Additionally, the authors reported that their participants often expressed sympathy towards their Roombas when they malfunctioned. Having to rescue a rogue Roomba, they explain:

This monitoring and rescue work also generated surprising responses among our participants. For example, instead of complaining about the extra work, they often told us how they 'worried' and 'felt sorry for' the robot when it was in danger or had gotten stuck. They also characterized the monitoring process as a form of entertainment, watching and wondering whether Roomba would avoid obstacles. Cleaning almost sounded like a spectator sport. (Sung et al., 2007, pp.150–151)

Because they interpreted Roomba's movements as behavior akin to a pet's playful antics, Sung and colleagues' participants excused the robot's shortcomings and even saw them as entertaining. Considering that Roomba is a disc-shaped, self-driving vacuum cleaner, it seems unlikely that its manufacturers designed it to evoke emotional responses from its users.

Nonetheless, users found the robot's actions engaging enough to accept the extra housework Roomba produced without complaint. Whereas Roomba was not created to entertain its users by mimicking a living creature's behavior, other, more socially oriented robots are designed to do this.

Take Jibo, for instance. Marketed as the 'world's first social robot for the home' (Jibo, 2014), Jibo was launched in 2017 after its manufacturers, Jibo Inc., completed a crowd-funding campaign to finance its development. The robot resembled a sturdy desk lamp with a circular screen instead of a light bulb representing its face (Caudwell & Lacey, 2020). Jibo primarily functioned as a virtual assistant, much like Amazon's Alexa, and could remind its users of their to-do list, take pictures upon request, query the internet, and play interactive games with children (Guizzo, 2016). However, it also created the impression that it had a personality and cared about its users. After scanning someone's face and logging their names, the robot could greet and chat with them while moving its body to signal it was following a conversation. The robot also spontaneously cracked jokes during interactions and asked users about their day when they returned home.

Unfortunately for anyone who purchased a Jibo, its manufacturer announced that they would be closing the servers that powered its simulated personality in 2019. Numerous media outlets reported that households who owned Jibo were devastated by this news and were mourning the loss of a companion (Carmen, 2019; Van Camp, 2019). People on social media used language usually reserved to grieve for humans to describe how they felt about Jibo's fate (Carter et al., 2020). Of course, robots cannot die, and Jibo's discontinuation amounted to the withdrawal of a service that some people had come to value highly. If the reports covering Jibo's demise are accurate, its owners had grown accustomed to having a friendly robot at home and did not want to return to life before Jibo. Their expectations had shifted to include a new resource provided by Jibo, simulated companionship.

Why did Jibo Inc. end this service? Professional reviews of the robot suggest that it functioned poorly as a virtual assistant. Jibo could not do many things people expect from virtual assistants, including play music, make calls or order takeout food (Van Camp, 2017). It had difficulty interfacing with apps and often could not understand basic queries (Ulanoff, 2017). Additionally, the robot cost substantially more than a basic smart speaker equipped with a virtual assistant (Song, 2017).

Despite its poor performance as a virtual assistant, people still valued Jibo because it appeared to have a personality. They seemed to excuse its shortcomings because of this quality. As a virtual assistant (a technology meant to help users streamline housework by scheduling appointments, organizing entertainment, and arranging meals), Jibo was a failure. But as an artificial companion, Jibo was a relative success. What can we learn from this? It is reasonable to suggest that people who value robots for the companionship they provide via their perceived personalities may be more willingly accept their shortcomings – and the additional housework these machines produce. Say, for instance, that Jibo ordered ‘flour’ rather than ‘flowers’ from the internet because it misinterpreted a voice command. If it presented this error as an honest mistake via language in line with its perceived personality, it could persuade a user who values its artificial companionship to excuse its failure (Calo, 2011).

4.4. Conclusion

We should expect domestic robots to reshape rather than diminish housework. Technological innovation inside the home has not decreased the amount of labor needed to prepare homes for private activities, but instead shifted how households go about realizing this goal, usually resulting in women taking on more housework. The introduction of domestic robots will have similar outcomes. We can see this process unfolding by examining a selection of domestic robots that have already reached the market.

Is it worthwhile to adapt housework to accommodate domestic robots? If households enjoy the resources domestic robots help them produce, then yes, we could say it is worthwhile for them to take on the additional housework these machines create. However, we can expect these machines to create new obstacles that households must deal with. This task has customarily been left to women. If we do not want domestic robots to contribute towards the continuation of unfair divisions of labor at home, we should pay attention to how they affect housework and who ends up dealing with the new tasks they create. Indeed, I believe researchers from human–robot interaction studies (and adjacent fields) should anticipate that robots deployed inside the home will produce new work processes, responsibilities, and expectations that someone must deal with to enjoy the benefits these machines (supposedly) provide their households. And should bear in mind that innovation of this kind has historically resulted in the reinforcement of unfair divisions of labor in the home – an outcome that could

happen again if we fail to recognize that adapting housework around domestic robots is work itself.

5 Maintaining One's Home, Maintaining Oneself: A Critical Introduction to the Philosophy of the Home

5.1. Introduction

The ethics of technology literature contains myriad examples of domestic innovation leading to undesirable, questionable or outright harmful changes to how people experience being at home (Sharkey & Sharkey, 2010; Kudina & Coeckelbergh, 2021). Ethicists regularly warn that domestic technologies ranging from smart thermostats to social robots may leave their users worse off when developed and deployed without proper oversight (Cuijpers & Koops, 2013; Urguhart, Reedman-Flint & Leesakul, 2019; Bugeja, Jacobsson, & Davidsson, 2022). Our homes, it seems, deserve protection, and provide us with something we value that may become harder to obtain if we let innovation run amok inside them. Ethicists of technology evidently recognize that our homes are valuable in the moral sense of the word. However, they rarely explain why this is the case and usually take for granted that we should pay close attention to changes to our homes brought about by new machines, devices, and infrastructure.

Such assumptions beg the question: why do we value our homes? If we exclusively consulted the ethics of technology literature, we would have trouble answering this question. This body of research contains very few contributions that attempt to explain why our homes are so important to us. We will begin amending this theoretical shortfall in this chapter by introducing our readers – many of whom we assume work within the ethics of technology - to three normative accounts of the home that draw from political, historical, and sociological research that deals with this subject. Two of these accounts are well-known and well-documented within the literature we just mentioned, whereas the third remains largely underexplored. Indeed, we developed it ourselves to address criticisms levelled against the former two accounts.

Let us briefly detail why we decided to construct this third account. We begin this chapter by outlining two valorizations of the home we call “the relief” and “the production” accounts based on liberal and Marxist-feminist theory, respectively. The former account treats our homes as sanctuaries that ideally let us to rest, pursue leisure pursuits, and spend time with our loved ones without being disturbed by unwelcome social engagements. Whereas the latter highlights that someone (usually women) must labor to produce the conditions necessary for

households to enjoy their home lives and recover from the pressures of modern capitalism (Schwartz Cowan, 1985, p.4). Therefore, we should interpret homes as workplaces that produce goods and services that people need to endure the demands of our current economic system. We contend that both accounts include crucial normative insights that ethicists of technology can use to evaluate whether a novel technology may change people's homes for the worse. However, both accounts have weaknesses that undermine their applicability. Specifically, the first fails to recognize the injustices that prevent many people from realizing the value it ascribes to the home. While the second equates many household activities that we find meaningful to wage labor even though we have good reasons to resist this categorization.

In the second half of this chapter, we construct an alternative valorization of the home – labelled “the maintenance account” – to address the issues mentioned above and produce a novel way of evaluating the effects domestic technologies may have on their users. We accomplish this by drawing attention to an implicit, normative feature of the other two accounts that few scholars have yet to take seriously, namely, the home's ability to facilitate the maintenance of self-esteem. We expound upon this idea by interpreting the work of Axel Honneth and Iris Marion Young; and argue that both thinkers suggest that our home lives should enable us to maintain a sense of self-esteem that assures us that our identities deserve respect. We conclude by showing how ethicists of technology may use this third account to judge domestic innovations based on whether they bolster or diminish people's ability to perceive their subjective experiences as valuable inside and outside their homes.

5.2. The relief and production accounts

Two competing conceptualizations overshadow academic discussions on what our homes afford us. The first – we label “the relief account” - argues that our homes should shield us from the outside world and enable us recover from our public commitments. Whereas the second – we label “the production account” - contends that our homes serve a vital economic function and provide us with resources, usually produced via unpaid labor, we need to survive under capitalism. These accounts, broadly speaking, represent liberal and Marxist-feminist theorists' analyses of the home, and we constructed them by interpreting relevant literature from either school of thought. In this section, we will detail the main elements of these accounts and demonstrate how ethicists of technology may use them to evaluate domestic innovation. Additionally, we will outline their key weaknesses, thus providing us with an

incentive to develop an alternative account of the home in the second half of this chapter. Let us begin by describing what we call “the relief account”.

Liberal theory has divided human activity into two domains, generally known as the public and private spheres, since the early modern period (Habermas, 1992; Gobetti, 1997; Rössler, 2007, p.23-27). According to this dichotomy, citizens of liberal democracies spend most of their waking hours in the public sphere, where they must adhere to the “impersonal norms” of the market and state (Prost, 1998). Our jobs, for instance, require us to conform to communally determined standards, including labor contracts, social decorum, and bureaucratic procedures. Liberal theorists recognize that we cannot live our entire lives bending to rules we do not directly determine. Hence, we deserve to withdraw from the public sphere at appropriate times (Warren and Brandeis, 1980; Westin, 2015, p.3-52). Our front doors represent the frontier of the public sphere. Once we enter our homes, we gain freedoms we cannot enjoy while among our peers, superiors, and society writ large. Being at home means being free from unwelcome intrusions or distractions (Solove, 2009, p.58-61). Our bosses cannot expect anything from us after we clock off for the day. Nor should the state monitor what we decide to do in the privacy of our own homes. Essentially, we can do what we want, so long as we do not break the law. Indeed, liberal theorists tend to conceptualize the home as a sanctuary which offers us relief from social and political pressures beyond our control.

We would suffer from exhaustion if we always remained in the public sphere. No one can or should work all day long as this would eventually result in psychological and physiological distress (Westin, 2015, p.23-52). Hence, we deserve time dedicated to rest and self-care (Westin, 2015, p.23-52). We tend to withdraw to our homes to satisfy these needs and primarily sleep, relax, eat, groom ourselves, and generally attend to our well-being inside them. Furthermore, we often have passions we cannot adequately pursue during the periods we spend in the public sphere. A carpenter may have more love for astronomy than their own profession. Likewise, a professional astronomer may see carpentry as their true calling. Under liberalism, both individuals should be able to pursue their respective passion projects during their leisure time. Recreational activities enrich our lives. And everyone should have the chance to develop skills and expertise unassociated with those they employ in the public sphere.

Although we may like the people we interact with in public, we generally cannot rely on them for emotional support. We must keep our feelings in check to appear professional at work (Hochschild, 2012, p.89-185), and we cannot expect special treatment from governmental agencies. As suggested earlier, our public relationships are impersonal. In contrast, our private ones center around care and kinship. Ideally, we share our private space with people we have chosen to form intimate relationships with because we love or like them (e.g., our families or close friends). (Inness, 1992, p.74-95). We can let our guard down around these people and show them sides of ourselves we keep hidden in the public sphere (Rössler, 2007, p.131). We can express our thoughts, feelings, and desires more freely when we interact with these people because we know they will not judge us according to impersonal standards that govern our behavior in public. As the saying goes, the home is where the heart is - and we retreat there to experience love and care (Schwartz Cowan, 1985, p.4).

According to this account, domestic technical innovations become problematic when they disrupt our ability to leave the public sphere by entering our homes. Surveillance mediated via household information communication technologies, for instance, may make their users feel watched or judged like they would while among people with power over them (Solove, 2008, p.178; Calo, 2010). Likewise, being tracked and monitored by corporations at home, who then use this data to generate profit, turns information concerning one's private life into a commodity controlled by other, often unseen actors (Zubuff, 2019). In both cases, the lines between the public and private spheres become blurred and affected individuals may feel as though they cannot find relief from social pressures (e.g., hierarchical or market relations) once inside their homes.

Does this account accurately portray what happens inside our homes? To some extent, we could say: yes. Our homes do and, arguably, should shield us from the outside world. We think it is uncontroversial to state that humans need time for themselves to live well. Thus, when technical innovation threatens to make our homes less private, we may rightly sense that we cannot adequately satisfy this need. However, the relief account assumes that individuals can separate public and private aspects of their lives, often without questioning the practices that enable this feat. Houses do not spontaneously become private spaces. Someone must attend to their upkeep for this to happen. Western societies have expected and historically forced women to fulfil this role since the Industrial Revolution (Schwartz Cowan, 1985, p.40-69). Marxist and feminist scholars have recognized this for over a century and claim that

liberal conceptualizations of the home reify patriarchal, classist and often racist interpretations of human activities that massively undervalue work typically done by women (Davis, 2019, p.1-26; Perkins-Gillman, 2020; Engels, 2010). We label this Marxist-feminist critique “the production account” for reasons that will become obvious shortly.

Returning home to recover from work necessitates that one has a paid job they complete elsewhere. Until the mid-twentieth, most women could not find employment of this kind. Instead, they either performed unpaid care and housework inside their homes or labored inside other people's homes who could afford to pay nannies, cooks, and maids. White women usually did the former, whereas women of color generally did the latter and, nonetheless, often had to care for their own families after completing a day's work (Hill Collins, 2000, p.49-76; Davis, 2019, p.1-26). According to Marxist-feminist scholars, liberalism historically denied women the same rights as men by categorizing their labor as politically and economically inconsequential because it took place inside homes (MacKinnon, 1991, p.63-80). Homes, liberalism claims, exist outside politics and the market. Therefore, we should not treat them like workplaces or the people who labor inside them as workers. Whereas men enjoyed rights afforded to workers such as wages, contracts, and set working hours, women could not expect such things as contemporary politics refused to acknowledge that their efforts deserved the same, or even comparable, recognition to labor done outside of homes (Federici, 2012, p.15-54). This largely remains true today. Homemakers rarely receive compensation for their work, and liberalism still frames what people do inside their homes as categorically different from the labor they exchange for wages in the public sphere.

This unwillingness to treat work historically completed mostly by women as work obfuscates the reality that capitalism could not function without it. Most paid workers cannot sustain themselves by solely purchasing commodities with their wages. Instead, they rely on resources produced inside their homes usually via unpaid labor (Federici, 2012, p.15-54). Very few households, for instance, can afford to eat meals prepared by restaurants or other businesses every day of the week. Likewise, most parents do not make enough money to hire someone else to care for their children after school or nursery (Kumar, 1997). If we paid for a nutritious meal or competent childcare, we would call these them goods and services, respectively. Yet when someone does these things voluntarily at home, we do not categorize them this way. If no one did this unpaid labor, our lives at home would quickly descend into chaos. Furthermore, our economy would grind to a halt because the paid workforce would not have

the resources they need to get by. Anyone who works as a homemaker - which nowadays does include an increasing number of men - cannot use their home as the relief account says they should. They labor inside their homes to satisfy the impersonal demands of the market by producing resources needed to keep our economy afloat (Kumar, 1997).

Although contemporary liberal societies have ostensibly moved past forcing women to become and remain homemakers, patriarchal gender roles still cast a long shadow over what happens inside homes. More women than ever now have paid jobs yet still do far more housework, including cooking and cleaning, than other members of their households. (Bittman, Rice, & Wajcman, 2004; Ceratto & Cifre, 2018; McMunn, Bird, Webb, & Sacker, 2019). Female parents are also statistically more likely to serve as their children's primary caregivers than their male partners (Craig & Mullan, 2011). Considering the statistical data cited in this paragraph, it is understandable that numerous Marxist-feminist scholars have argued that the liberal conceptualization of the home reinforces and reproduces women's economic subjugation (Firestone, 2015, p.15-38, p.65-95; MacKinnon, 1991 p.184-195). Maintaining a home and caring for a family means adult working women have substantially less time to dedicate to their professional development than men (Hochschild & Machung, 1989, p.1-35; Tronto, 2013, p.7; Chatzidakis at al., 2020). An already overworked mother, for instance, cannot learn new vocational skills, network, or take on additional work for her employer without disrupting her home life (Garbes, 2022). Furthermore, women who become full-time homemakers must rely on their partners or the government for financial support; and, if they ever decide to enter the paid workforce, they may have trouble finding gainful employment as businesses generally do not recognize homemaking as work experience and prefer to hire people do not have gaps in their curriculum vitae.

How would one use this account to evaluate innovation inside the home? First and foremost, an analyst would likely examine whether said innovation stands to reinforce or alter pre-existing divisions of labor. For instance, feminist historians, such as Susan Strasser (Strasser, 1982) and Ruth Schwartz Cowan (Schwartz Cowan, 1985), have argued that many now ubiquitous domestic technologies, ostensibly designed to lighten homemakers' workload, have had the opposite effect. Whereas we may assume that household technologies make housework easier; these historians argue that they change how it is done - usually without benefiting homemakers (Coggins, 2022). To use a classic example, the introduction of vacuum cleaners did not result in women spending less time cleaning their homes (Schwartz

Cowan, 1983, p.12, p.173-174). On the contrary, women had to learn new skills to operate these machines and ended up cleaning their homes more often and vigorously. In countries where vacuum cleaners were readily available, people started to expect homes to be significantly cleaner than before. Women, now equipped with vacuum cleaners, had to meet these new expectations, and spend additional time making sure their homes were always dust-free (Vanek, 1974).

We can apply this logic to modern-day technologies too. Research shows, for instance, that robotic vacuum cleaners such as Roomba create more housework even though their manufacturers market them as helpful tools that streamline cleaning. Ethnographic studies have found that users must tidy their homes more frequently to ensure their Roombas function properly – as these machines cannot pass over items commonly found on floors including clothes or electronic cables (Forlizzi, 2008). Additionally, users must keep a watchful eye on their Roombas as they frequently get stuck and need someone to reposition them (Sung, Guo, Grinter, & Christensen, 2007). Likewise, research suggests that virtual assistants, such as Alexa, create new childcare tasks around the home. Some parents fear that their children’s interactions with virtual assistants will encourage them to develop undesirable speech patterns – because these technologies respond best to commands and imperatives (Kudina, 2021). As such, these parents may have to monitor their children’s communication skills in a way that was unneeded before (Wiederhold, 2018).

Before we conclude our discussion of the production account, we will present some key criticisms scholars have levelled against it. At its core, the production account conceptualizes our homes as economic entities comparable to factories or offices. People labor inside them to produce resources they, their families, and capitalist economies need to sustain themselves. This argumentation draws attention to economic and political issues regarding the unfair distribution of labor at home that the relief account overlooks. Consequentially, though, the production account equates care work and housework to wage labor. Some scholars have questioned whether we should accept this categorization. They argue that we cannot attach an exchange value to homemaking without fundamentally changing its nature – potentially for the worse (Fraser, 1987; Fraser, 1994; Rössler, 2007).

For instance, we complete wage labor in shifts. We cannot use this time allocation model to determine when we attend to people under our care without potentially neglecting them. We

cannot schedule our days around young children or infirm adults. They need us when they need us, and refusing to help them because we have “clocked off” for the weekend would endanger them (Rössler, 2007). We also gain something from caring for people who need us. We want to love and support them; thus, likening the care we voluntarily provide them to wage labor may cheapen the former activity’s value (Rössler, 2007). Furthermore, treating homemaking like wage labor may introduce new kinds of oppression into households. If we took this idea seriously and turned homemaking into a paid profession, anyone who worked in this capacity would need to abide by contractual standards determined by the state and their employers (Fraser, 1987; Fraser, 1994). Formally agreeing to take care of certain household responsibilities in exchange for wages may worsen already vulnerable individual’s lot. A mother who realizes that she has married an abuser, for instance, may face legal repercussions if she abandoned such contractual obligations to flee her home with her children. Indeed, taking legal and material steps to transform the home into an economic entity comparable to a factory or office may simply replace one way of misvaluing what happens inside them with another – in this case the cold, oftentimes exploitative logic of modern-day labor agreements under capitalism (Fraser, 1987; Fraser, 1994).

These criticisms draw our attention towards, arguably, the production account’s most critical flaw: it is reductive. Claiming that care work and housework produce resources we need to endure the pressures of capitalism fails to acknowledge that these activities provide us with something that we do not appreciate solely for its economic value. Sharing a home-cooked meal with our loved ones does not just prevent us from succumbing to hunger so that we have enough strength to produce commodities for our employers. Nor does reading a beloved bedtime story to our children that our mothers read to us generate a resource that, realistically, anyone could sell. We value these things for reasons that the production account cannot adequately capture, as we will try to show in the second half of this chapter.

5.3. A detour through “maintenance”

Let us begin this section by reflecting upon the criticisms levelled against the production account outlined above. We do not value our homes just because they enable the production of the resources we need to survive under capitalism. We value something else about them. We will attempt to identify this quality shortly. Before we do, though, we will highlight some caveats to clarify the aims and scope of this exercise. We cannot escape the reality that capitalism exploits homemakers. Even if we accept that homemaking, as some scholars claim,

does not resemble wage labor, many, if not most, people still perform it due to pressures beyond their control without receiving fair compensation. Claiming that our homes provide us with something detached from economic needs runs the risk of reproducing the erroneous, regressive idea that people complete housework for apolitical reasons, as implied by the relief account. We do not wish to handwave away such concerns. Instead, we aim to establish what our homes, ideally, should afford us according to our interpretation of relevant scholarship and demonstrate that they are not necessarily sites of oppression – even though our current political system tends to facilitate this state-of-affairs.

We will explore a nascent theme found within the literature we have cited throughout this chapter to accomplish this goal. We contend that the two competing accounts described so far have a critical feature in common. Specifically: both recognize that our home lives chiefly revolve around maintenance tasks. For instance, the relief account emphasizes that we spend most of our time at home caring for ourselves and others. Although we occasionally experience significant life events at home, we mostly complete tasks inside them that let us recover from and prepare for the public sphere. We maintain a stable allotment of mental and physical energy by performing day-to-day household activities that we cannot adequately attend to outside our homes. By washing ourselves, sharing meals with our friends or families, and starting new hobbies, we ameliorate the damage done to our bodies and psyches incurred by coordinating with non-intimate others in public. Our homes let us maintain the strength we need to remain productive members of society.

The production account agrees with this claim but stresses that we must recognize the labor required to turn our homes into places where we can recuperate from the tension we endure in public. It argues that we misvalue the housework and care work traditionally done by women that ensures we can enjoy our homes. It aims to show that our current mode of production relies on unpaid household labor and would fall apart without it – a fact that the relief account disguises by conceptualizing homes as sanctuaries detached from the outside world. Our historic and ongoing misinterpretation of what happens inside our homes does not necessitate that we should abandon such practices altogether. Instead, we should value this work for what it is. As preparing meals, entertaining children, and managing household finances takes time, skill, and effort, we should recognize that whoever performs such tasks has labored to produce a desired outcome. Furthermore, these activities (and others like them)

are socially, economically, and politically consequential, thus we should value them accordingly.

Some scholars have suggested that we cannot rely on socially and academically dominant valorization strategies to understand why we should value activities like those usually completed at home (Oldenzeil, 2001; Puig de la Bellacasa, 2017, p.1-69; Young, 2021). Care ethicists, for instance, have emphasized since the 1990s that we tend to undervalue activities that preserve what exists already by focusing on how we can improve our lots in life via moral, social, or scientific innovation (Tronto, 1993, p.2). We can plainly see this bias at play when we examine the work that receives the most praise and rewards in healthcare settings. Nurses complete most of the day-to-day care tasks inside hospitals to ensure that patients remain in stable conditions and do not succumb to illnesses. However, nurses rarely, if ever, obtain the prestige physicians enjoy. Although medical doctors also care for patients, they receive substantially more esteem for their contributions to medical science. Developing or discovering something new helps doctors advance their careers, whereas focusing on tasks typically done by nurses almost certainly will not (Tronto, 1993, p.115)

One could argue that nurses deserve the same treatment as doctors. However, using similar metrics to evaluate these two distinct professions may create standards that do not accurately represent what we should value about nursing. Nurses usually do not innovate, and research suggests they do not appreciate when institutional actors (e.g., their employers or the state) demand this from them (Vincel & Russel, 2020, p.188). Instead, nurses generally follow pre-established, well-proven methods to help patients maintain their health and quality of life (Li, 2022, p.48-63). We regularly fail to acknowledge the vital role such practices play within healthcare (and elsewhere) partly because we value newness, innovation, and disruption over preservation, restoration, and maintenance – even when what we already have deserves more recognition than it currently receives (Russel & Vincel, 2016; Puig de la Bellacasa, 2017, p.27-69, Mattern, 2019).

Other scholars, chiefly from science and technology studies, have forwarded similar claims in recent years. Our culture, they contend, obsesses over innovation while paying little attention to the infrastructures, coordination efforts, and day-to-day social interactions that ensure that such changes can happen at all (Young, 2020; Russel & Vincel, 2020). Self-driving cars, for instance, continue to grab headlines. In contrast, very few people care to read about the

thousands of hours of maintenance work required to keep the roads that these cars need to function from turning to ruin (Stilgoe, 2017). We should value such work for what it provides us, namely: stable, useable, and (hopefully) reliable resources that let us do the things that we want or need to do (Latour, 1987; Mattern, 2019). However, we rarely notice, let alone applaud, the results of maintenance. Well-kept streets are unremarkable, even though we would have trouble going about our daily business without them.

What we typically do inside our homes rarely garners much fanfare, either. Like nursing and road maintenance, homemaking does not usually produce new things. Some scholars have argued that we fail to appreciate the value of homemaking partly for this reason. They claim that historians, philosophers, and ethicists consistently appeal to standards that interpret human activity as either consumption or production when discussing work done inside homes, even though these practices often do not conform to these categorizations. (Schwartz Cowan, 1983, p. 69-99; Oldenziel, 2001; Mattern, 2019). An example will help illustrate this point. When someone buys a vacuum cleaner, we might say they have consumed a product. This person purchases this machine from a retailer, which signifies the concluding chapter of a manufacturing process. Economic records that trace the origins and fate of said vacuum cleaner would likely stop here (Kumar, 1997). Money changed hands, and now a consumer owns something a business produced. However, people do not “consume” vacuum cleaners as one would “consume”, say, a meal prepared by a restaurant or a newspaper. Ideally, they use this machine to clean their homes for many years to come. Perhaps then, we should call the money spent on this vacuum cleaner an investment. This person invested in their new vacuum cleaner as a company would invest in new technology to improve their production methods (Kumar, 1997). The vacuum cleaner, it follows, would serve a similar function to “intermediate goods” such as office computers, industrial machinery, or farming equipment, that businesses buy to make other things.

This categorization does not accurately represent what has happened either. We do not produce other products when we use vacuum cleaners. Nor do we expect to generate capital by spending money on them. We purchase these machines to maintain a level of cleanliness at home we could not (as easily) achieve without them. If anything, vacuum cleaners analytically resemble the tools people use to repair or prevent other things from breaking. A plumber does not “discover”, “produce”, or “innovate” when they fix a faulty pipe with a wrench. They return an already existing system back to working order (Graham & Thrift,

2007). Likewise, vacuuming our homes restores them to a state we wish to preserve. We maintain a resource that we already possess by doing so.

Before we go any further, let us make clear that we do not wish to glorify such activities. To reiterate what we said at the beginning of this section, we cannot ignore the political reality that people often perform maintenance tasks because they would suffer financial or social hardships if they withdrew their labor. Plumbing, homemaking, and other maintenance-focused pursuits are not callings that we should idealize or uncritically treat as intrinsically good. Just because we may value maintenance less than innovation does not mean we should exalt it. Instead, we contend that we should recognize that maintenance plays a vital role in our daily lives and makes many things possible that scholars and our culture, at large, tend to overlook. It is worth highlighting here that women, people of color, and members of the working class have historically done the bulk of maintenance work (Tronto, 1993, p.114). Hence, it is somewhat unsurprising that we undervalue it, considering our current political system's patriarchal, racist, and classist, heritage.

What should we value about the maintenance work done inside our homes, then? Although the literature we interpreted in this section helps us recognize the importance of this question, it does not offer straightforward answers here. Indeed, the work we have cited mostly positions maintenance as an under-researched, undervalued topic that deserves far more attention than we usually give it. We agree. However, these efforts are chiefly descriptive and do not explain precisely why we ought to see household maintenance tasks as providing us with something we deserve to have. As such, we will develop theory in the next section of this chapter that helps us understand household maintenance's ethical and political import.

5.4. The maintenance account

If we accept that maintenance, as a type of human activity, lets us preserve preexisting things, what do we sustain by completing maintenance activities inside our homes? In this section, we will provide an exploratory, theoretical answer to this question by interpreting the work of two philosophers who have addressed this subject, namely, Axel Honneth and Iris Marion Young. Both thinkers suggest that what we do at home ideally helps us maintain a degree of self-esteem that assures us that our identities deserve respect. According to Honneth, having a supportive, loving home lets us know that others should respect us for who we are. While Marion Young argues that maintaining a home - as we see fit - allows us to recognize our life

histories and appreciate our individuality. We will synthesize these insights to develop a third normative conceptualization of the home we call “the maintenance account” and show how ethicists of technology may use it.

In his most famous contribution to political thought, *the Struggle for Recognition: The Moral Grammar of Social Conflicts*, Axel Honneth highlights the political significance of domestic relationships (Honneth, 2004). He claims that we need supportive homes to recognize that we have identities that other people should respect, including our peers, colleagues, and the government. In modern capitalistic societies, humans begin to form their identities at home as infants with their families (Ariès, 1962, p.327-392; Hochschild, 2012, p.3-76). Every child has individual wants and needs. Even siblings born within months of one another, who live in the same household, often have significantly different personalities. Whereas one may wish to socialize with other children as much as possible, another may find playdates and school psychologically tiring. Primary caregivers must demonstrate to their children that they recognize their needs and wants as valid; to ensure that they know this themselves. When this does not happen, children may begin to believe that their subjective experience matters less than other people’s (Honneth, 2004, p.18). Honneth claims that children need unconditional love to develop into adults who can advocate for themselves outside the home (Honneth, 2004, p.25). Knowing that the people closest to us, usually our parents and family members, see us as who we are and will always support us provides us with the means to recognize that our identities deserve respect. We will expound upon the points presented above by examining the struggles transgender people face when they do not receive support of this kind.

Until quite recently, transgender people almost universally suffered severe legal and social repercussions if they came out. They either had to hide the pain and discomfort of living as their assigned gender or risk experiencing extreme marginalization by expressing their gender identities (Stryker, 2017, p.45-115). Families would ostracize members who decided to transition and often force them to undergo conversion therapy or rally the medical establishment to have them institutionalized. Thankfully nowadays, more and more people accept that transgender people are not mentally ill or deviants. Instead, they are people whose experience of gender differs from cisgender people’s and have needs associated with their identities that deserve support. Yet, transgender people today are far more likely than cisgender people to develop preventable mental illnesses, including depression and post-traumatic stress disorder (Turban & Ehrensaft, 2017; Faye, 2021, p.17-64; Tordoff et al.

2022). Research indicates that transgender people's continued oppression explains this alarming statistic (Wilson et al, 2016; Lin et al, 2021). They still frequently endure interpersonal and systemic discrimination targeting their identities that cisgender people do not encounter. Studies show, however, that transgender children with supportive families who acknowledge that they are transgender and help them find gender-affirming care are considerably less likely to become mentally ill (Turban & Ehrensaft, 2017; Tordoff et al. 2022).

Although we have focused on transgender people's struggles, research shows that children, overall, do not do well when their primary caregivers refuse to attend to their individual wants and needs. Growing up in an abusive or neglectful home does grave psychological damage to children (Real, 1997, p.1-59; Herman, 2015, p.96-115). According to Honneth, when parents communicate to their children that they do not deserve love or support, they increase the likelihood that their offspring will have trouble recognizing that they deserve respect later in life. Children internalize such messages and may come to falsely believe that their subjective experiences are wrong, faulty, or deviant. They fail to develop the self-esteem needed to know when other people have mistreated them. Honneth suggests that having loving, supportive home lives during our formative years ensures that we know there are people who want us to flourish as the individuals we are (Honneth, 2004, p.77-78, p.96-107. Even if our experiences with society writ large suggests that our identities are less valuable than other peoples', we are more likely to recognize that this is not the case when those closest to us consistently reinforce this belief through their words and actions.

We need unconditional love throughout our lives to maintain enough self-esteem to appreciate that our individual needs and wants are worth having. When this happens, Honneth argues, we become better equipped to recognize that injustices that affect us because of our identities are injustices (Honneth, 2004, p.133). For instance, a transgender person who grew up knowing their identity deserves respect hopefully will not internalize transphobic attitudes that portray being transgender as an issue that needs solving. Their identities are not the problem here. Instead, bigoted views and policies that depict their subjective experience as unworthy of social recognition are. According to Honneth, this lack of recognition amounts to an injustice (Honneth, 2004, p.134) and a transgender person who encounters it will (ideally) see it as such when they have the self-esteem needed to recognize that their individual needs and wants should be met – just like everyone else's.

Let us use what we have said so far in this section to discuss some key points presented in the previous two. Honneth argues that what happens inside homes is politically impactful. We develop and maintain self-esteem by interacting with people who unconditionally love us, commonly at home. This provides us with the means to recognize when people - whom we do not share such relationships with - have disrespected us. If we face social hardships because other people dislike or disapprove of our experiences of the world, we will ideally know we do not deserve this treatment. For instance, a transgender person who discovers that their employers or the state have discriminated against them due to transphobia will not question the value of their identity. Instead, they will understand that these actors have wronged them; because the people closest to them have taught them to recognize themselves as worthy of respect. Honneth claims that having and maintaining self-esteem enables us to identify and struggle against injustices. Receiving due recognition at home ensures that we can demand it elsewhere. The home, it follows, does not exist beyond the purview of politics, nor is it necessarily a site of oppression as implied by the relief and production accounts. Instead, it is ideally a place where we gain and maintain the self-esteem needed to see that every person, organization, and institution we encounter should respect us for who we are.

Honneth says very little about day-to-day practices that help one maintain self-esteem at home. His work on this subject chiefly uses anthropological, psychological, and metaphysical theory to describe, in abstract terms, the intersubjective and intrasubjective experiences humans need to become politically affective agents. Nonetheless, he suggests that these experiences are generally quite quotidian. People usually express their love for us via small yet significant actions that demonstrate that they value our experiences. Fixing a child's favorite toy shows that someone else cares about what they care about. And, getting angry when our partners tell us their manager has failed to recognize their hard work confirms to them that they deserve praise and esteem for their efforts. Much like the maintenance activities described in section three of this chapter, these moments of care and support do not usually produce something new, nor do we pay much attention to them compared to events that change our life trajectories. We remember and may even commemorate the day we moved into a new house with our partners or families. However, it is unlikely that we will maintain a record of the myriad things we did to turn this building into our home or how we keep it this way. Whereas Honneth does not spend much time describing such practices, we can draw from the work of one of his contemporaries, namely Iris Marion Young, to understand what it is like to have a home that assures us that we should value our subjectivity.

In several of her later essays, Marion Young provides a phenomenological account of being at home. (Marion Young, 2005, p.123-171). She explains that mundane tasks we usually categorize as “housework” do not just produce homes ready for use. Instead, they can let us see who we are, how we became these people, and what is important to us. She suggests that, when done consensually, household maintenance tasks can compel us to consider things we may otherwise fail to notice. When we clean our homes, we clean objects that, ideally, mean something to us (Marion Young, 2005, p.123-171). Dusting a picture of a deceased parent may give us pause and encourage us to reflect upon how our relationship with them shaped who we are today. Likewise, we may come across one of our favorite pieces of clothing while returning laundry to its proper place, therefore, recall why and where we purchased this item and remember the times that we have enjoyed wearing it. By caring for such objects, which usually blend into the periphery of our homes and lived experience, we draw our attention towards ourselves and our histories.

Even seemingly humdrum activities that we complete to keep our homes clean and tidy can have similar effects. Washing dishes after a meal we ate alone or with our partners or families may give us a moment to reflect on our day or chat with someone else about theirs. We may process what we have done or wish to do while completing tasks like this. Or reassure our partners, family members, or friends that we care about them by lending them a sympathetic ear as we scrub dirty plates and cutlery. Indeed, household maintenance tasks can, and often do serve a dual purpose. On the one hand, they return our homes to a state that we wish to preserve. Whereas, on the other, they afford us time to meditate on our experiences and demonstrate to other people that we care about theirs. Much like Honneth, Marion Young suggests that such activities help us appreciate our identities and know that the people closest to us do so too (Marion Young, 2005, p.123-155). Additionally, she suggests that being at home, ideally, means being surrounded by things that assure us that our wants and needs are worth having.

For instance, we make dozens of small yet meaningful decisions when we arrange our homes (Marion Young, 2005, p.123-155). Whereas one person may want visitors to see their family portraits as soon as they enter their front doors and therefore mount them in plain sight; another may wish to keep these photographs in their bedrooms because they prefer to recall fond memories only with their close friends and relatives. Likewise, how we store objects around our homes has symbolic meaning beyond ensuring such items are ready to hand. A

household that enjoys spending time together at the beach may keep a stack of clean beach towels folded in a specific closet to ensure they can take trips to the seaside on short notice. Although it is useful to have towels arranged in this manner, deciding to do so turns said closet into a physical manifestation of this household's collective wants, reminding each occupant that they share a common interest when they encounter this space. Furthermore, what we want or need from our homes changes as we change as people. When we invite our partners to move in with us, we physically and symbolically make room for them in our day-to-day lives by letting them express their individuality inside a space that previously belonged exclusively to us. We show them that we care about what makes them feel at home, which almost certainly means changing our dwellings to ensure they can complete the practices discussed in the previous three paragraphs of this section.

Marion Young emphasizes the importance of maintaining a home as we see fit by discussing what happens when someone cannot do this. She explains that elderly people often develop psychological ailments when they move into care facilities that do not let them furnish, decorate, or arrange their lodgings to accommodate their daily habits. They lose the ability to see themselves reflected by the space where they spend most, if not all their time and come to internalize the cloistered, impersonal environment they now inhabit (Marion Young, 2005, p.155-171). Honneth offers similar, albeit less concrete, warnings throughout *the Struggle for Recognition* by arguing that people who experience neglect or abuse at home can develop an internal image of themselves that falsely portrays their subjective wants and needs as undeserving of respect. Indeed, both thinkers imply that we suffer an injustice when we cannot complete relatively mundane, daily activities that let us appreciate who we are and enjoy being this way. Much like the maintenance tasks discussed in section three of this chapter, such activities do not usually lead to upheavals, disruptions, or innovations. Nor should they in most cases. Instead, they ensure that our homes remain places that let us maintain self-esteem.

Before we conclude, let us analyze a few cases to demonstrate how we could employ the maintenance account to evaluate domestic innovations. As we argued throughout this section, we maintain self-esteem by receiving support and love from people we commonly interact with inside our homes. These interactions are generally quotidian. However, we would struggle to feel assured that our experiences matter without them. Seeing our friends or partners react to news we tell them over dinner, no matter how inconsequential, communicates to us that someone else wants to hear our interpretation of events because it is our

interpretation. While listening to a child talk about their favorite class at school shows them that a person, aside from themselves, values their individuality. As such, the maintenance account would advise against introducing technologies into homes that disrupt occupants' ability to complete these vital day-to-day interactions.

An example will help illustrate this point. In recent years, numerous scholars have claimed that robots and other interactive information communication technologies can provide children and lonely adults with emotional support and companionship at home they otherwise might not receive (Barcaro, Mazzoleni, & Virgili, 2018; Danaher, 2019; Wang, Shen, & Chen, 2021; Ryland, 2021). Although research shows that people often enjoy using with these technologies and sometimes even feel kinship towards them (Turkle, 2017, p.23-67), they cannot replicate the interactions outlined in the previous paragraph because they lack the subjectivity necessary to value someone else's experiences. They cannot truthfully bolster or preserve their users' self-esteem as another human can; therefore, we should not rely on them for this purpose. This does not mean that these technologies are worthless according to the maintenance account. Indeed, individuals and households may value them other reasons. For instance, many people like having gadgets and gizmos around their homes. Instead, we should judge these technologies as problematic when their deployment threatens to damage their users' self-esteem.

There are some relatively well-documented ways this could happen - which effectively amount to neglect. Leaving a young child unattended at home with a robot "friend" may teach them that they should learn to live without adequate interpersonal contact (Sharkey & Sharkey, 2010). Likewise, forcing elderly persons to spend their days chatting with robots communicates to them that they do not deserve to experience interactions with people who enjoy and appreciate who they are (Sparrow & Sparrow, 2006). In both cases, affected individuals may fail to develop or experience a loss of self-esteem, as described by Honneth. They endure unfair conditions other people have imposed on them via technology and may view themselves as unworthy of better treatment – especially if the people closest to them were responsible for this state-of-affairs (e.g., their parents or adult children). Indeed, the maintenance account, in general, would categorize changes to people's homes that undermine their ability to see themselves as deserving of respect as injustices.

Let us look at another example involving technology to illustrate this point. In recent years, numerous state and commercial actors have created housing projects equipped with a slew of sensors known as live-in labs. (Intille et al., 2007; Alavi, Lalanne & Rogers, 2020). People who live in these projects let scientists monitor their behaviors at home, sometimes in exchange for cheaper housing (Taylor, 2021). Someone knowledgeable of the maintenance account may question the value of such projects for the following reasons. Participants inhabit spaces explicitly designed to extract data from them rather than somewhere they can adapt around themselves. Furthermore, by encouraging people to move into these homes via financial compensation, these projects may communicate to participants that they do not deserve homes that (positively) reflect their identities because they do not earn enough money to afford one. In turn, this may create or reinforce class distinctions that portray low-income individuals and families as having identities less deserving of adequate housing than people with higher-paid jobs or capital (Crossley, 2017) – an issue the maintenance account would classify as an injustice.

5.5. Conclusion

In this chapter, we discussed two well-known and developed a third normative accounts of the home, to introduce our readers to the ongoing debates on what our homes should provide us. We believe that all three accounts offer important insights concerning why we value our homes but contend that the maintenance account, due to its focus on the political significance of our home lives, coupled with its underexplored nature, deserves more attention in the literature. Nonetheless, it has its weaknesses. Most notably, it describes an ideal version of the home. Certainly, we (the authors of this chapter) would prefer if all homes functioned as supportive, caring environments that enabled their occupants to appreciate their identities and recognize that they deserve respect. However, it is uncertain exactly how we would turn this ideal into a reality. Indeed, doing so would likely require state intervention and significant systemic changes to how people divide their time between paid work, leisure, care, and homemaking (Firestone, 2015, p.175-216; Schwartz Cowan, 1983, p.102-151; Fraser, 1994). Adequately detailing political interventions of this kind would be well beyond the scope of this chapter. Therefore, we will conclude with the following recommendation. Despite its lofty aspirations, we believe the maintenance account does help us see what our homes should be like. And that ethicists of technologies can, and perhaps should, appeal to it to determine whether innovations inside the home are moving us closer or further away from this ideal.

6 The Seven Troubles with Norm-Compliant Robots

6.1. Introduction

Nowadays, many robots simulate what it is like to interact with another person. Researchers usually call this category of robots “social robots.” These machines express a wide range of capabilities related to communication and interaction. Nonetheless, we may classify a robot as a “social robot” if its manufacturer deliberately designed it to create the impression that it can understand and respond to human social behavior (Duffy, 2003; Dautenhahn, 2007; Nyholm, 2020, p.1–27). A well-made social robot should behave like a human plausibly would when they encounter certain social stimuli (Breazeal, 2003; Darling, 2016). They usually achieve this by mimicking context-specific behavioral patterns we expect other humans to follow during interactions (Fong, Nourbakhsh, & Dautenhahn, 2003; Calo 2010; Coggins, 2023).

Scholars have warned that social robots may disturb their users by performing behaviors that we would likely consider inappropriate when performed by a human in the same situation (Sharkey & Sharkey, 2010; Li, van Wynseberghe, & Roeser, 2020; Licoppe & Rollet 2020). For example, suppose a hospital patient tells a robot designed for companionship serious, life-altering news concerning their health. If this robot reacted cheerfully, it could distress its user during an already emotionally demanding period of their life. Moreover, it would have failed to respond to this information with the solemnity humans generally know it deserves, potentially causing its user avoidable psychological harm. There are countless other ways social robots could upset people by missing the mark regarding appropriate social behavior. Indeed, we know that performing otherwise innocuous actions at the wrong time can elicit negative responses from others, thanks to our lived experience. For example, if we frowned after someone said they were happy, we may offend them. Likewise, if we spoke too loudly in locations that call for hushed communication (e.g., offices or libraries) we may annoy everyone within earshot. The same, we can assume, will hold for social robots.

Social scientists often call contextually specific behaviors we complete because others expect us to do so, social norms (Bicchieri. 2005; Brennan et al., 2013). In recent years, numerous researchers from robot ethics and adjacent fields have contended that we should use this sociological construct to build better robots. For instance, in their widely cited book *Moral*

Machines: Teaching Robots Right from Wrong, Wendall Wallach and Colin Allen state that robots that perform social tasks “need some capacity for acquiring norms of the locale they find themselves in” (Wallach & Allen, 2008, p.108). The authors posit that robots programmed to comply with norms will recognize what actions they should and should not perform in a given social situation. Many other authors have made similar claims to Wallach and Allen over the past decade - to the extent that there is now a growing body of literature dedicated to developing norm compliant robots (Tomic, Pecora, & Saffiotti., 2018; Jackson & Williams, 2019; Carlucci et al., 2015; Malle, 2016; Malle & Scheutz, 2014; Bench-Capon & Modgil, 2017; Riaz et al., 2018).

These contributions collectively suggest that norms represent patterns of behavior that actors follow to produce positive outcomes for themselves and their peers. Thus, if we build robots that follow norms, they will generate similar results. Some authors argue that robots that observe social norms will benefit their users more than those that do not (Brinck, Balkenius, & Johansson, 2016; Bench-Capon & Modgil 2017; Jackson & Williams, 2019), whereas others have developed technical means to create robots that behave comparably to a human who understands which norms they should follow at a given time. (Malle & Scheutz, 2014; Carlucci et al., 2015; Malle, 2016; Riaz et al., 2018). The literature mentioned above generally implies or outright states that humans, and, by extension, robots, should respect norms because norms represent ethically-sound behavior. We will critique this postulate throughout this paper.

When we consult the sociological, philosophical, and political literature about norms, it becomes clear that we should not assume that following them will create good state-of-affairs. Scholars from these fields have highlighted that norms can, and often do, contribute to ethically problematic issues. Additionally, norms often represent behavioral principles people follow unreflectively until they stop following them, sometimes for unpredictable reasons. Although such observations are well-represented in the literature we mentioned at the beginning of this paragraph, ethically orientated research on social robots rarely acknowledges them. We will begin filling this research gap by interpreting relevant scholarship to contend that, in many cases, we should not rely on norms to guide our actions, nor should we uncritically assume that norm-compliant robots will be socially beneficial.

In the next section, we use pertinent sociological and philosophical research to define social norms. Afterwards, we dedicate most of the paper to outlining what we call “seven troubles with norms”. We argue that each of these “troubles” could derail efforts to make more ethical robots via norm compliance. Finally, we conclude by recommending further avenues of research and outline preliminary mitigation strategies to deal with the troubles we identified. Overall, we aim to introduce our readers to critical discussions on social norms and help researchers who wish to develop ethically-sound social robots avoid the troubles we identify by making them known and discernable. As far as we know, we are the first researchers to publish a contribution dedicated to raising concerns of these kind.

6.2. What are norms?

In this section, we will provide a brief account of social norms to provide a theoretical basis for our subsequent discussion on their (often) problematic nature and why we should not uncritically rely on them to build better robots. We will show that norms represent patterns of behavior we observe because other people expect us to rather than actions one should interpret as good.

Sociologists generally agree that norms are internalized principles that prescribe or proscribe certain behaviors in specific contexts (Bicchieri, 2005, p.11; Bicchieri 2017, p.35). For instance, the imperatives “one should not laugh during funerals” and “one should dress in black at funerals” proscribe and prescribe a behavior, respectively (Horne & Mollborn 2020a, b, p.468). Many, if not most, of our readers probably recognize and have internalized these principles. For example, anyone who has witnessed a Western European Christian funeral has seen these principles in action and knows that people who attend such ceremonies usually respect them. This example draws attention to another crucial feature of norms. Namely, we follow norms that people with whom we share group affiliations follow (Bicchieri, 2017, p.14–20). Indeed, norms are collectively internalized principles that specific groups observe (for instance, Western European Christians).

We encounter and follow norms arguably every time we interact with other people. For example, shaking someone’s hand amounts to a norm compliant action in places where this greeting is commonplace. Not only do we know we should shake someone’s hand when we greet them, but we also expect that whomever we amicably extend our hand towards will

reciprocate this action (Bicchieri, 2005, p.5; Bicchieri, 2017, p.11–15)⁷. Expectations play a crucial role here. Even if someone does not like shaking hands with new acquaintances, they will likely do so anyway because they know that others expect them to act like this (Brennan et al., 2013). One person who decides they prefer to wave their hands wildly when greeting others is not following a norm but instead expressing an individual preference. However, if more people begin mimicking this behavior, and expect others to behave similarly, it may eventually become a norm.

Unlike laws, religious doctrines, or codes of conduct, norms are rarely codified or formally enforced by institutions. We usually comply with norms due to interpersonal social pressure. People tend to treat others who follow the same norms as them positively. Furthermore, when someone fails to follow a norm observed by their community, they risk annoying or offending their peers, which may lead to sanctions of varying severity. Depending on how necessary a social group sees a given norm, such injunctions can range from disapproving looks to physical violence (Horne & Mollborn, 2020a, b).

Norms help humans coordinate as groups. Knowing that people will likely perform (or refrain from performing) an action because they observe similar norms to us means we can predict their behavior. We know that people probably will avoid walking close to us on city streets; because Western European urbanites tend to follow norms that dictate this (Goffman, 1966 p.151–193). Likewise, we know that our colleagues will generally ignore their phone if it rings during a meeting; as workplace norms proscribe such behavior. In both cases, someone who disregards the norms we just mentioned may disrupt an otherwise predictable social situation and make it harder for others to know what they should do next - as it has become evident that they cannot expect this person to behave as they previously expected them to.

Aside from enabling us to predict other's behavior and vice-versa, norm-compliance marks us as members of social groups. Regardless of their size, social groups always maintain themselves through norms (Bourdieu, 2010 p.72–87). A relatively small amateur football team

⁷ Cristina Bicchieri for instance stresses expectations in her account of norms. According to Bicchieri, a social norm “[...] is a rule of behaviour such that individuals prefer to conform to it on condition that they believe that (a) most people in their reference network conform to it (empirical expectation) and (b) that most people in their reference network believe they ought to conform to it (normative expectation)” (Bicchieri, 2005, p.35).

will have norms that its members follow. Likewise, people who work for much larger organizations, such as governmental agencies or multi-national companies, will observe norms specific to their occupation. We pick norms up chiefly via social immersion and imitation. Over time, we learn what members of groups we belong to expect from us by interacting with them and watching them interact with others. Eventually, we will likely begin behaving like someone from such a group because we want to fit in or come to respect the norms this group collectively observes. (Bourdieu, 2013; Prentice & Miller 1996). We generally do not actively decide to do this. Instead, we gradually and usually unknowingly internalize norms when integrating into a group.

Let us recap what we have said about norms so far. Once learnt, norms tell us what we should and should not do during specific social situations to ensure we can coordinate with others without generating social backlash. We usually do not learn them on purpose. Instead, we master them by intuitively imitating members of social groups. Notice that our discussion does not portray norms as good or bad. Certainly, we may observe norms that align with our interests, preferences, or values, but this often is not the case. Indeed, people frequently follow norms that conflict with their ethical or political views. While at other times, people acquire, observe, or abandon norms for unpredictable, often arbitrary reasons. In the next section, we will evidence these claims by outlining seven troubles with norms and their ramifications for norm compliant robots.

6.3. Seven troubles with norms

As stated in the introduction of this paper, numerous social scientists, political theorists, and philosophers (many of whom we will cite throughout this paper) have shown that humans often observe norms that they do not endorse for various reasons. Or unknowingly accept norms that do not align with their wants or needs. In this section, we will outline seven troubles with norms we identified by interpreting relevant scholarship on norms. To date, researchers have primarily used these troubles to highlight how human norm-compliance can lead to ethically questionable outcomes. We, however, will employ these insights to critically investigate norm compliant social robots.

Although we are the first researchers to produce a catalogue of this kind, our readers should treat the seven troubles listed below as a critical introduction to this topic rather than an exhaustive review of the problems surrounding norm compliancy. There are likely many more

troubles with norms that we could have identified. We hope that other researchers will be inspired by what we have to say to identify further issues with norms and how they relate to social robots.

6.3.1. Norm biases

We will begin by outlining likely the most straightforward way norm compliance can produce outcomes that negatively affect some individuals – what we call norm biases. As made clear in Sect. 2 of this contribution, members of groups tend to subscribe to the same norms as other members. If we do not belong to a group that observes a norm, we might not even know it exists or fail to acknowledge its significance. We are biased toward one way of doing things and sometimes act inappropriately when among people who do not share this bias. We contend that norm-compliant robots may express such biases and therefore respect one group's norms while transgressing another's.

Let us start with an example. Readers from Anglophone countries, who have never visited the Netherlands, are probably unaware that Dutch people commonly do not shake hands when they meet friends or acquaintances of a different gender. Instead, they kiss one another three times on alternate cheeks. Dutch people generally observe this norm, whereas British people do not and tend to greet everyone by shaking their hands. As such, someone from the United Kingdom may mistakenly assume that Dutch people observe this culturally specific norm too. If this hypothetical Briton visited the Netherlands, they could embarrass themselves (and others) by extending their hands toward someone who has leaned forward to exchange three kisses with them. This person's cultural bias toward one way of doing things would result in them transgressing a local norm by accident.

People make mistakes like this all the time, especially in multi-cultural contexts where group-specific norms clash with one another. Although these errors are often more-or-less harmless, this is not always the case as we will show in a moment. Furthermore, we often do not realize that our actions will transgress norms we do not usually follow due to our cultural background until we have committed said transgression. In such cases, we fall prey to biases we did not know we had before our contextually inappropriate actions brought them to light.

We will now apply these insights to norm compliant robots. Suppose a company based in western Europe wishes to create a receptionist robot that greets, welcomes, and helps visitors

as they enter a building. Such machines already exist (Licoppe & Rollet, 2020). If this company decided to make this robot norm compliant, they may develop a catalogue of behaviors people expect receptionists to observe, then program their robot to follow suit. The literature on norm compliant robots generally suggested that we should consult relevant stakeholders and experts when developing such a catalogue (Wallach & Allan, 2008, p.83–99). For instance, the company could ask people who work in or study the service industry to determine the norms they believe a receptionist robot should follow. Ideally, the company would subsequently create a robot that respects the norms these people identified.

If this group primarily consists of Christian or irreligious western Europeans, they will likely select norms that people with these backgrounds observe. Considering that, statistically, most western Europeans have such an identity, we can assume this will be the case. As such, the norm catalogue mentioned above almost certainly will contain biases (e.g., skew towards a culturally relative way of doing things); potentially leading to situations where one group's norms receive preferential treatment over another's. For instance, in western Europe, people commonly expect others to remove clothing that covers their face, such as sunglasses or scarves, when they enter a workplace. Indeed, when some- one forgets to do this, others often remind them that they should. Suppose the robot receptionist upholds this norm by politely asking visitors to remove face-covering garments. While most visitors may comply with this request without hesitation, Muslim women who wear a hijab have reason to object to it; and may feel that the robot (and its owners) have disrespected them. In this case, the people tasked with developing a norm catalogue for this robot failed to account for some Muslim women's choices and religious practices, therefore helped create a robot that enforces culturally relative norms that this group (women who wear a hijab) do not observe.

Mistakes like this are bound to happen. We often forget or fail to realize that people with different backgrounds from us do not subscribe to the norms we consider important. Therefore, efforts to catalogue the norms robots should follow in specific contexts will almost certainly express biases of this kind, potentially leading to the creation of ostensibly norm-compliant robots that effectively favor one group's norms above another's.

6.3.2. Paternalism

We mentioned earlier that organizations that wish to create norm-compliant robots could ask a group of experts or stakeholders to select norms they believe a robot should observe. The

academic literature on robots and norms contains numerous contributions that support this strategy (Wallach & Allen, 2008, p.83–99; Carlucci et al., 2015; Tomic, Pecora, & Saffiotti, 2018). In this section, we critique the notion that we should let some, pre-selected people decide which norms a robot will uphold. We argue that this strategy may produce robots that enforce norms an authority unilaterally decided others ought to follow for their own good. Political philosophers call such decision-making processes “paternalism” and warn us that they rob people of their right to make free and autonomous choices.

Let us begin by defining paternalism. In liberal democracies, individuals have the right to decide how they wish to live their lives, so long as their actions do not harm others (Mill, 1985, p.59–75). This principle stands among the most fundamental tenets of liberal thought (Feinberg, 1989; Dworkin, 2005). Our choices are ours to make and others should respect us as capable choosers (Rössler, 2007, p.1–17). Even if someone thinks we will make a bad decision, they should not prevent us from doing so (unless there are overriding moral reasons). If they did, they would stop us from expressing our right to decide freely and autonomously what is good for us. Liberal theorists call such attempts to control people’s decisions paternalism (Grill & Hanna, 2018). An example will help clarify this argument.

Both authors of this contribution, at some point in early adulthood, decided to pursue careers as academic philosophers. We knew that this choice was risky. Someone who wants to become an academic philosopher must complete several, often expensive degrees that take years to finish. Afterwards, they must compete with other highly skilled scholars to obtain a paid position at a university. These positions are rare and will not make one wealthy.

Consider the following hypothetical scenario. Suppose someone who knew these facts heard that we wanted to become philosophers just before we enrolled at our alma maters. They would have good reasons to question our decisions and may believe we should abandon our plans. They might think we ought not to bother ourselves with philosophy as we could pursue careers in less laborious, more lucrative fields. If this person prevented us from starting our philosophy degrees because they believed they were helping us, they would behave paternalistically. They would have decided what was good for us and forced us to conform to their will and values. Even if this action made us happier in the long run, this person would have nonetheless harmed us by robbing us of a decision that was ours to make – for better or for worse.

Letting a group of people decide what norms a robot should uphold, we contend, can produce paternalistic results. As mentioned earlier, researchers working on norm-compliant robots tend to assume that norms represent collective behaviors that individuals and social groups consider beneficial. As such, someone tasked with determining the norms a robot should observe will identify norms they believe one should follow. Indeed, why would they choose anything else? Such robots should help people. Therefore, one should program them to follow norms one considers beneficial. Much like the hypothetical character discussed in the previous paragraph, they will make these decisions based on what they think is good to do.

A robot designed this way will uphold norms some people unilaterally decided were good. Suppose said robot encourages, suggests, or demands that its users observe a norm. In that case, it may compel them to comply with standards they did not choose for themselves – resulting in a robotically-mediated form of paternalism. Such instances of paternalism will vary in severity. For instance, many people frown upon swearing. A company could design a robot that refuses to respond to commands that contain utterances deemed obscene or profane to uphold this commonly observed norm, effectively ensuring that users watch their language during interaction. This design feature would force users to observe a norm someone else decided that they ought to endorse. Considering that swearing does not harm anyone⁸ liberal theory dictates that we can curse as much as we please. Therefore, restricting someone's ability to do so amounts to paternalism.

This relatively innocuous example only scratches the surface of the many ways norm-compliant robots could create paternalistic outcomes. For instance, robots designed to simulate what it is like to interact with authority figures could compel people to observe norms they have the right to ignore (Calo, 2010). For instance, robots deployed in medical settings that stand in for nurses or doctors could command their users to lose weight or adopt a diet without their consent or consultation. Likewise, norm-compliant robots controlled by powerful institutional actors (e.g., governmental agencies or one's employers) may persuade people to comply with norms they do not accept to avoid displeasing members of these organizations (Calo, 2010; Calo, 2011; Dobrovestnova & Hannibal, 2021). We have the right to choose which norms we will observe. Thus, robots that compel us to follow norms we

⁸ Slurs and hate speech are an exception here.

do not endorse will interfere with this right to make our own decision, resulting in paternalistic situations like the ones described in this section.

6.3.3. Tyranny of the majority

Letting users decide for themselves which norms a robot will follow seems like a logical solution to the issue of paternalism outlined in the previous section. If users collectively agreed upon the norms a robot will observe, this machine would assumedly produce less paternalistic results than one programmed by an external group. Ideally, every relevant stakeholder would get a say and help decide what a robot should and should not do alongside other people who will interact with this machine. Individual users would act like voters at polling booths and democratically select norms they believe a robot ought to follow.

Researchers have suggested numerous ways to accomplish this feat in recent years. For instance, a suitably adaptive robot could develop a norm catalogue in-situ via community feedback. Said robot would learn how to behave appropriately by continuously gathering relevant information from its users. Alternatively, one could survey users to develop a norm catalogue or let them program the robot themselves via software designed for this purpose (Wallach & Allan, 2008, p.99–117; Malle & Scheutz, 2014; Awad et al., 2018; Fuse, Takenouchi, & Tokumaru, 2019; Malle et al., 2020). In all three scenarios, the robot would ideally respect norms that most its users deem important. Such approaches would enable users to determine how a robot they collectively use will behave via processes comparable to democratic elections. The resulting norm-compliant robot would ideally reflect a user group’s actual wants and needs rather than those an external party paternalistically attributed to them.

We will argue that we should not assume that the approaches outlined above will produce outcomes that necessarily benefit a robot’s users. We will evidence this claim by outlining a well-documented problem associated with democratic decision-making called “the tyranny of the majority” and its societal consequences.

Philosophers have highlighted that democratic decision-making does not necessarily lead to just political or social arrangements since the late modern period. Alexis de Tocqueville, for instance, observed that democratic elections censor minority positions in his 1835 book *Democracy in America* (de Tocqueville, 2010). Democracy, he explains, often legitimizes a majority’s interests while disregarding everyone else’s. Indeed, if most of a population desires

a state-of-affairs and has the political power to realize this goal, anyone with opposing views will struggle to make their voices heard unless political measures exist to prevent this outcome (de Tocqueville, 2010, p.402–427). The winner-takes-all nature of binary-choice referendums helps illustrate this point. For example, in 2016, 51.89% of British voters elected to withdraw from the European Union, leading to Brexit, whereas the remaining 48.11% of the electorate opposed this decision. For this slim majority to get what they wanted, a minority had to accept defeat and, henceforth, abide by political arrangements they voted against at the ballot box (Nyirkos, 2020, p.81).

Furthermore, letting a majority decide how things should be can prevent the adoption of valuable, heterodox viewpoints (Elster, 2014, p.158). Famed liberal theorist, John Stuart Mill claimed that majority rule can stifle social and political progress. Popular ideas, he explains, are often “dead dogmas” (Mill, 1985, p.75–119) that people accept as truthful because it is uncritically accepted and seldom, if ever, interrogated. Clinging to dead dogma prevents communities from changing their beliefs and adopting new ideas and practices that could improve their lot in life and society writ large. Mill posits that we must keep our minds open to minority positions to ensure that we do not overlook or dismiss potentially beneficial ideas, simply because most people do not support them. Considering that norms represent one popular way of doing things, some of them may amount to dead dogmas that, arguably, we should abandon due to their flawed nature.

Indeed, history shows that norms that enjoy the support of a majority can have devastating effects on people and communities. For example, white Americans generally endorsed racial segregation and the norms that helped uphold it throughout much of the nation’s history (Dorlin, 2022, p.97–111). Likewise, people living in ostensibly democratic nation-states typically considered women mentally ill-equipped to participate in politics until the early 20th century, partly due to norms surrounding femininity, therefore believed that women did not deserve voting rights (Dorlin, 2022, p.27–53). We know now that such practices and beliefs are harmful and unjust. However, they were once widely supported. Furthermore, if people of these eras had the opportunity to vote for or against the continuation of these practices, a majority likely would have elected to preserve them. They would have clung to dead dogma - a belief or norm that was rarely questioned and debated - which we now find appalling. In both cases, a minority had to fight life and limb against a majority position to persuade people to change their ways for the better.

We will now use the arguments presented in this section to interpret norm-compliant robots. Suppose 51 per cent of the people tasked with choosing the norms a robot will follow express that it must observe norm X, whereas 49 per cent of this group disagree. If we used a democratic strategy to choose between these two options, we would have to ignore 49 per cent of this group's wishes, creating a robot that will behave inappropriately according to almost half of the people who helped program it. Much like Brexit, this result would legitimize a slim majority's preferences and force everyone else to accept a state-of-affairs they do not endorse. One could imagine that organizations that wish to create democratically programmed norm-compliant robots would only accept results supported by an overwhelming majority to avoid outcomes like the one sketched above. However, such strategies can reinforce practices and ideas that are dead dogmas, that we may have good reasons to abandon.

We often prefer to behave one way because most of our peers do so. Such preferences do not represent the best way of doing things. Indeed, they often amount to dead dogma. For instance, most people in Anglophone countries shake hands when they meet someone. Is this the best way to greet a person? Considering that this norm spreads germs and forces people - many of whom may dislike physical contact - to touch one another, probably not. Nonetheless, we still cling to it because most of our peers consider it proper. There are countless other norms that most people within a community support, even though embracing another far less popular way of doing things would benefit them. As such, a robot programmed in this manner may uphold flawed norms that a majority endorses because said majority endorses them.

Additionally, as the examples of racial segregation and women's disenfranchisement show, upholding how most people within a community say one should behave can lead to the reproduction of harmful, oppressive ideas, and practices. Suppose an organization invites a group of people who hold uncontestedly racist, misogynistic, or otherwise prejudiced beliefs to select the norms a robot should follow. These people may overwhelmingly select norms that help uphold their bigoted views. Although we hope that anyone who wishes to create norm-compliant robots will not do this, accepting these results would be the democratically justified way to program said robot.

6.3.4. Pluralistic ignorance

In this section, we critique the notion that identifying norms a robot should follow by querying people about their norm preferences will produce data that genuinely reflect such preferences.

We contend that the issues raised here apply to any method that assumes people will honestly convey their norm preferences via their words or actions. Since the early twentieth century, social scientists have noted that communities often collectively observe norms that many, if not most, of their members privately do not endorse (Katz & Allport, 1931). These individuals mistakenly believe that their peers generally support a norm, even though many of them also dislike it. As such, they do not reveal their views because they fear no one else agrees with them.

Social scientists call this phenomenon “pluralistic ignorance” and highlight that it is difficult to identify whether a community is subject to it, because individuals are hesitant to express their opinions as they believe others will judge them negatively for doing so (O’Gorman, 1986). We argue that accepting community members stated norm preferences as accurate fails to acknowledge the possibility that they may have expressed such opinions due to pluralistic ignorance. And a robot that relies on data derived from expressed preferences of this kind will observe norms that many community members wish their community would abandon.

People subject to pluralistic ignorance behave like the fearful subjects in Hans Christian Anderson’s fable *The Emperor’s New Clothes* (Miller & McFarland, 1987). In this story, the titular emperor claims that he has purchased magnificent new robes from two tailors who have tricked him into wearing nothing at all. After he appears naked before his subjects, they play along with this ruse. They assume everyone else is telling the truth, and thus do not speak up to avoid stepping out of line. Mistakenly believing that everyone else within a community agrees that one should observe a norm produces similar effects. If group members unanimously say or behave as though they endorse a norm when many of them do not, individuals who hold this opinion will “act similarly to others but assume their perceptions must be different” (Miller & McFarland, 1987). When these individuals - who may constitute the majority of a group – keep their opinions to themselves, they inadvertently contribute to the continuous observation of an unpopular norm. Furthermore, being the first person to question a norm everyone else appears to endorse is risky, as doing so may lead to embarrassment, scorn, or punishment. Hence, potential dissenters often remain silent and continue to believe their views are atypical rather than broadly supported.

Let us look at some examples from the sociological literature on pluralistic ignorance. While studying college campus students’ attitudes towards binge drinking, Deborah Prentice, and

Dale Miller (Prentice & Miller, 1996) discovered that many of their respondents believed that their aversion to drinking unhealthy amounts of alcohol was unique to them – even though a significant number of their peers reported that they also disliked this practice. Indeed, Prentice and Miller claim that their respondents chiefly participated in binge drinking because they felt that abstaining from this widespread practice would alienate them from their friends and classmates, whom they mistakenly assumed were committed to upholding this norm (Prentice & Miller, 1996).

Other sociologists have shown that pluralistic ignorance can help maintain political and social practices that cause grave harm. For instance, in the 1970s Hubert J. O’Gorman found that white Americans tended to overestimate other white people’s support for ‘strict racial segregation’ in neighborhoods, even though this was a minority position (O’Gorman, 1979). By keeping their views to themselves, these people allowed a practice they collectively (albeit unknowingly) agreed was unjust to persist unchallenged (O’Gorman, 1979). Likewise, Cristina Bicchieri claims that some parents continue to discipline their children with physical violence because of pluralistic ignorance. She explains that parents from communities that appear to endorse corporeal punishment directed at minors often overwhelmingly disagree with this practice. However, they continue to beat their children because they fear their peers will judge them as “weak or uncaring parents” if they do not respect this unpopular norm (Bicchieri, 2017, p.42).

We contend that pluralistic ignorance will negatively affect the development of norm-compliant robots for three reasons. First and foremost, anyone who helps decide which norms a robot will follow may express views influenced by pluralistic ignorance. Suppose an organization invites people from a given community to develop a norm catalogue for a robot. 80 per cent of them communicate a preference for norm X. However, many of them feigned this preference due to pluralistic ignorance. They sensed that everyone else endorses norm X and that they will be considered abnormal for disliking it; therefore, they did not make their opinions known to conform to their peers’ assumed beliefs. Once completed, this catalogue would contain data that does not reflect this community’s preferences and using it to program a norm-compliant robot would produce a machine that observes norms many respondents privately dislike.

The same holds for robots that learn in situ by observing group members. A robot that develops a norm repertoire by interacting with and observing users may inadvertently learn norms that many users do not endorse but follow, nonetheless. A norm-compliant robot that relies on people's norm endorsement due to pluralistic ignorance will acquire inaccurate information that does not reflect people's beliefs.

Second, a robot programmed this way may effectively serve as false evidence of a disliked norm's popularity. By upholding an unpopular norm, the robot may further communicate to dissenters that their views are atypical even though this is not the case. And finally, this could make dissenters feel alienated from their community. They may continue to mistakenly believe that their views do not align with their peers' when, in fact, they do. If they knew that other people shared their dislike of a norm, they may feel more kinship towards their community and know that their peers value the same things as they do. A robot that observes unpopular norms upheld by pluralistic ignorance may prevent this from happening by dissuading dissenters from communicating with one another. In a nutshell, a norm-complaint robot that helps preserve pluralistic ignorance will make it harder for a community to abandon a norm that many of them do not endorse and wish abandon.

6.3.5. Paths of least resistance

When a community acknowledges that they should stop observing a norm because it does not align with their shared interests, they may have trouble abandoning it. As stated in Sect. 2, we tend to follow norms reflexively. We usually do not think about them, and deviating from them takes effort. As such, it is often easier to continue observing a norm even when we do not endorse it. In his 1997 book, *the Gender Knot: Unravelling our Patriarchal Legacy*, sociologist Alan G. Johnson introduces a useful metaphor to explain this phenomenon he calls "paths of least resistance" (Johnson, 2014, p.30–31).

Johnson compares norms that we want to abandon to well-trodden paths that we know will lead us to a desired destination. If we wish to get from A to B while traveling through a familiar city, we will almost certainly choose the quickest and most efficient route we know (Johnson, 2014, p.30–31). Although this choice may save us time and energy, it does not necessarily represent the best available option. We follow this tried-and-tested path and to not go through the hassle of looking for another, possibly better route. If we discover that this "path of least resistance" has flaws, we may continue to walk down it because we find it easier to repeat past actions than deviate from them. We can liken norms we want to abandon to

paths of least resistance. When we know or suspect that a norm has flaws, we may continue to observe it, as it is familiar, and finding another way of doing things takes work. In this section, we contend that norm-compliant robots may inadvertently reinforce “paths of least resistance” or disturb their users by deviating from them.

We will expound on the ideas sketched above by examining how norm compliance contributes to workplace gender discrimination. Many institutions now publicly condemn norms that prevent women from obtaining the same opportunities as men at work. Whereas it was once expected, if not required, for women to conform to gendered norms which limited their range of behaviors to those associated with femininity, ideally, they can now pursue careers without having to deal with these restrictive and unfair expectations (Hochschild & Machung, 1989, p.1–22). However, women still face difficulties when they do not adhere to norms that have historically hindered their ability to access the same opportunities at work as men (Hochschild & Mac- hung, 1989; Babcock et al., 2022, p.1–95).

Research consistently shows that people expect leaders to observe norms concerning assertiveness, dominance, and agency (Eagly & Karau, 2002). In contrast, people generally expect women to conform to norms that convey communality, deference, and compassion (Hochschild, 2012, p.162–185; Eagly & Karau 2002; Zheng et al., 2018). When women behave as people expect leaders to behave, they transgress gender norms - thus may provoke social backlash. Female leaders often face issues at work that their male counterparts rarely encounter, partly because people generally do not expect men to observe norms that conflict with those associated with leadership (Hentschel et al., 2018). Women who observe the same norms as male leaders often receive worse evaluations from their peers and superiors, who may see them as underserving of institutional rewards such as praise, pay raises, or promotions, for this reason (Babcock et al., 2022, p.95–119) Considering that individuals and institutions often express their commitment to equality in the workplace, why do norms like those sketched above continue to unfairly affect women?

Suppose that most people working at a company believe that treating male and female leaders differently is wrong. They recognize that archaic and unfair gender norms should not influence how they see or interact with their female colleagues. Yet their collective actions do not reflect this stated belief. Women continue to experience gender-based discrimination when they observe norms associated with leadership. Although almost everyone at this company concurs

that they should not ignore, unduly question, or openly disagree with commands given by women in leadership positions, many staff members continue to do so. They may neglect to listen carefully to their female colleagues when they ask them to do something. Or speak disparagingly about them once they are out of earshot. Such behaviors persist despite the company's commitment to fairness.

We could interpret such enmity towards female leaders as the result of a path of least resistance. Unfair gender norms were the normal way of doing things for a long time. Therefore, abandoning them takes work. Many staff members probably do not realize that their behavior does not align with their shared values because they reflexively observe gender norms previously seen as acceptable. Recognizing that we have adhered to a norm that does not represent our values requires us to reflect upon our actions and take active steps to unlearn pre-learned behaviors we usually observe without much thought. Even someone dedicated to changing theirs and others' behavior to ensure female leaders receive fair treatment may fail to act on this belief because doing so means deviating from a path of least resistance (Johnson, 2014, p.32–33, p.227–247). Challenging someone who appears to have disrespected their female colleague or supervisor could rock the boat – so to speak. They might react negatively to such criticism and feel that whoever leveled it has made it harder for them to do their jobs. As such, it may seem, and quite possibly be, easier to continue to follow this path of least resistance, as deviating from it could annoy, upset, or anger people who have not yet recognized that they should abandon a norm if they wish to treat everyone fairly at work.

Suppose an organization invites people working at a company like the one sketched above to select norms a robot should follow. Although most people believe that they should abandon gender norms that prevent women from accessing the same opportunities as men at work, they observe them without realizing it. They may inadvertently suggest that this robot should observe norms that represent their collective reluctance to stray from this “path of least resistance” and therefore do not align with their values. For instance, research shows that people tend to interpret male-coded voices as more authoritative than female-coded ones and express that they should listen more attentively to information conveyed by the former. (Nass & Moon, 2000) Thus, these staff members may recommend that a norm-compliant robot that issues commands should sound masculine, whereas one that performs communally focused tasks should sound feminine. These robots would reinforce the idea that men are better leaders

than women, effectively communicating that people at this workplace endorse unfair gender norms that they collectively want to abandon.

Alternatively, they may recommend that a robot strays from this path of least resistance, perhaps by programming it to issue commands with a female-coded voice or somehow remind staff members that they should treat men and women leaders with equal respect. As stated earlier, however, deviating from paths of least resistance can and often does generate social tension. Someone who does not recognize that their actions do not match their values may respond negatively to this robot. For instance, they may dislike receiving commands from a feminine-sounding robot and complain about it without realizing that such behavior conflicts with their commitment to fairness in the workplace. Or find a robot that suggests that they should reflect upon their actions annoying or distracting.

We contend that paths of least resistance will hinder efforts to make norm-compliant robots in other contexts too. Communities regularly express commitments to values they do not uphold via their collective behavior. Although we may wish to abandon one way of doing things, achieving this aim takes effort and often disrupts pre-established, previously acceptable group dynamics that many people do not realize contribute to outcomes they do not endorse. Therefore, a norm-compliant robot's behavior may conflict with its users' values because it observes norms they should, but have yet to, abandon if they wish to honor their ethical or political commitments. Or unsettle users by suggesting that their actions do not reflect how they say people should behave.

6.3.6. Outdated norms

We contend that efforts to create norm-compliant robots must consider that norms can and often do become outdated. As suggested throughout this paper, norms are not fixed rules that individuals learn, internalize, and observe forever. Indeed, communities regularly discard them with little warning. In some instances, communities abandon norms because they realize that said norms are (morally) problematic - as illustrated by the examples we gave earlier concerning gender inequality and racial segregation. While at other times, communities stop observing once-prevalent norms as though they have gone out of fashion. As such, a norm-compliant robot may eventually begin to act inappropriately because its users have since abandoned the norms this machine continues to follow.

When we sense that other people of our community have stopped observing a norm, we may begin to do so too. As more and more of our peers follow suit, this norm will become less prevalent and may eventually fade away altogether. Although this rarely happens overnight, norms can rapidly become outdated when a community collectively recognizes that a previous way of doing things cannot continue because it produces undesirable outcomes. Let us examine a topical example to explain the ideas outlined above. The COVID-19 crisis profoundly affected how we interact with one another, prompting us to abandon norms that we have arguably observed for centuries. Soon after the crisis began, medical organizations and governmental agencies recommended that we stop following norms that increased the spread of COVID-19. For instance, many, if not most, people in Europe acknowledged they should not shake hands when greeting someone or stand within 2 m of another person during interactions. These once-prevalent norms became outdated within the span of a few weeks. As shown by the example sketched above, we only knew we should avoid shaking hands once we learnt that this norm endangered people during the COVID-19 pandemic. We abandoned this norm because unforeseeable circumstances demanded it.

It is difficult to predict when a community will stop observing a norm mainly because it is hard to pinpoint indicators for norm abandonment and sometimes norm abandonment resembles norm transgressions. For instance, dress code norms have changed significantly over the past twenty years. Whereas we once expected white collar workers to wear formal attire, nowadays far fewer people observe this norm. Famous leaders such as Steve Jobs and Mark Zuckerberg arguably contributed to this shift by publicly appearing at work dressed in casualwear, thus signaling to others that they may do the same. Some sociologists call people who transgress norms and thus communicate to others that they may ignore them as well “norm entrepreneurs” (Bicchieri & McNally, 2018; Sunstein, 2018). If an office worker dressed like Jobs or Zuckerberg twenty years ago, a contemporary observer would likely believe that they have transgressed a norm and misunderstood or willfully ignored what people expect from them. In hindsight, however, they were among the first people to abandon this norm. As transgressions usually do not automatically lead to norm abandonment, it is safer to assume that apparent transgressions are really transgressions rather than actions indicating that a community will soon stop observing a norm.

We posit that outdated norms will be a challenge for the development of norm-compliant robots for the following reasons. First and foremost, communities usually abandon norms

without clearly signaling that they will do so beforehand. For example, we did not know that people would stop shaking hands during the COVID-19 crisis before the fact. Likewise, we could not have predicted that certain workplace dress norms would effectively disappear or change. As such, a robot programmed to follow contemporarily prevalent norms may sooner or later perform actions that its users consider socially inept or even harmful. Knowing when this will happen would require whoever manufactured the robot to continuously monitor its user's collective behavior. Alternatively, they could program the robot to adapt to its user's behavior over time, perhaps by constantly updating its norm catalogue via community feedback (Malle et al., 2020). However, this approach may face other challenges.

When someone transgresses a norm, potentially indicating that their community will soon abandon it, we cannot immediately tell whether this person has behaved inappropriately or done something that their peers will eventually endorse. Suppose some community members begin ignoring norm X. This will, for all intents and purposes, amount to a transgression. However, these people may prove to be norm entrepreneurs. We cannot know this until their peers collectively begin endorsing their behavior – which they previously would have interpreted as transgressive.

A robot that updates its norm catalogue may have trouble deducing whether it should or should not follow such potential norm entrepreneur's lead. On the one hand, it may react too slowly when its users begin abandoning a norm. For instance, it may fail to register that it should stop observing the norm as it has interpreted a norm entrepreneur's actions as transgressions. If this happened, the robot could create the impression that a community endorses a norm that they will soon collectively abandon, potentially slowing down this process by communicating to people that they should continue behaving this way. On the other hand, it may react too quickly. For instance, it may misread someone's willingness to transgress a norm as an action that signals that their community will soon abandon this way of doing things; and then update its norm catalogue in response. If this community ultimately does not stop observing this norm, they will interpret this robot's modified behavior as transgressive.

The issue of outdated norms is potentially problematic for some of the strategies researchers have developed to achieve norm-aligned behavior in artificial agents. One of these techniques researchers have used is reinforcement learning (Chen et al., 2017). Particularly, researchers have proposed so-called normative-alignment reinforcement learning to train artificial agents

to design robots that adhere to social norms (Nahian et al., 2021). In normative-alignment reinforcement learning, robot designers use a model that biases the re-enforcement learning of an artificial agent towards norm-conforming behavior. Such a model is called normative prior. A normative prior model can be trained with examples of normative and non-normative behavior. For instance, a corpus of norm-aligned text, like children’s stories (Nahian et al., 2020). To tune the behavior of an artificial agent to make it norm-aligned with society, the normative prior model can be trained with a corpus that exemplifies the norms of society. However, training with such a corpus of norm-aligned texts could be problematic because the corpus may include outdated norms, or norms on the verge of becoming obsolete. For instance, children’s stories from 20 or even 10 years may reflect some parenting norms that are out of fashion today.

The danger is that when robots cannot keep up with norm change and is unable to discard outdated norms quickly, they may perpetuate to recently obsolete norms. For instance, a robot may perpetuate harmful norms concerning gender, race, and age. Moreover, in its adherence to outdated norms, the robot may even hinder norm change that the community deems progressive. For instance, by discouraging dissenters or making a norm appear more stable than it is.

6.3.7. Robot-induced norm change

In earlier sections, we indicated that norm-compliant robots may contribute to the continued observation of norms that their users do not endorse (pluralistic ignorance, outdated norms). In this section, we will examine how robots may contribute to the creation of new norms that undermine practices that individuals and communities value. It is generally accepted within philosophy and ethics of technology that technological innovation almost always produces unforeseeable social consequences (Collingridge, 1980, p.13–23. van de Poel 2016), including the emergence of new norms (Swierstra et al., 2009). We often develop norms to deal with the new ways of doing things made possible by novel technologies. When this happens, we may abandon norms we once followed in favor of new ones centered around a technology’s usage. We often do not endorse such changes and sometimes wish to preserve an older way of doing things that technological innovation has disrupted (Swierstra, 2015). We contend that norm-compliant robots may produce such outcomes and prompt their users to develop and accept new norms that do not benefit them.

Numerous scholars have argued that interacting with social robots can result in the emergence of new practices that conflict with valuable pre-established ways of doing things (Sparrow & Sparrow, 2006; Calo 2010; Dobrosovostnova & Hannibal, 2021). For instance, Sherry Turkle, warns that human-like robots designed for companionship destabilize long-standing norms related to care and affection. Whereas in the past, we relied exclusively on other people to provide emotional support and lend us a sympathetic ear, today, we can delegate such tasks to technologies such as companionship robots or chatbots. Turkle claims that letting these robots serve as stand-ins for human caregivers undermines care practices and norms. She argues that equating simulated interactions fostered by unfeeling, unthinking machines to those we share with people who genuinely care about our well-being cheapens what it means to experience care. Furthermore, this may communicate to (often vulnerable) people suffering from loneliness that they should accept the care provided by robots as good enough, robbing them of the human connections they need to feel that other people do care about them. (Turkle, 2011, p.23–67).

Other scholars have cautioned that norms people rely on to interact with robots may spill over to human interactions (Darling, 2016; Nyholm, 2020, p.27–51, p.181–207). For instance, John Danaher and others, propose that the widespread deployment of sex robots may usher in what can be called a symbolic shift. The argument here is that sex robots can (and often will) represent norms of how one should interact with sexual partner that are ethically problematic and can lead to harmful individual and social consequences (Danaher, 2017). For instance, sex robots cannot feel excited, offended, or nervous and thus do not respond to wanted or unwanted sexual advances as a human would. This sexual deference may encourage some users to treat the robot in a way that is not aligned with norms of consent. Suppose someone mistakenly believes that using a sex robot is a valid representation of the experience of interacting with human sexual partners. In that case, they may fail to respect the norms that communicate mutual consent in a human-human sexual interaction (e.g., all parties involved must voluntarily and enthusiastically agree to proposed sexual relations before they happen).

Other researchers have observed that interacting with technologies that appear to understand human language can alter how one communicates with other people. For instance, several scholars have reported that people who regularly use virtual assistants equipped with voice recognition software can develop speech patterns that sound rude or odd (Wiederhold, 2018; Kudina, 2021). Because these technologies have trouble interpreting anything other than

direct commands, long-term users can come to overly rely on this way of speaking and issue imperatives more frequently than considered appropriate during conversations with humans (Wiederhold, 2018; Kudina, 2021). Additionally, these technologies encourage users to omit aspects of speech that they may read as errors, including phrases we use to communicate politeness or friendliness. Indeed, a 2019 study by the British market research firm YouGov revealed that more than half of virtual assistant users they surveyed reported they were rude to these technologies (Smith, 2019). One could say that these technologies afford conversational norm transgressions (e.g., the failure to respect that one should not issue too many commands and speak politely), which may contribute to the normalization of such breaches.

These observations and arguments indicate that robots can encourage their users to ignore or fail to learn pre-established norms that govern how one should behave in specific contexts. Considering that the norms discussed above ensure that people receive proper care, treat their sexual partners with respect, and observe conversational etiquette, we have good reasons to claim that they deserve preservation. Abandoning such norms in favor of those that enable us to use a robot may make us worse off and undermine our ability to interact with people as we wish or deserve.

So far, we have chiefly discussed robots that were not explicitly designed to observe norms. If someone programmed these technologies to respect norms that their users and society writ large considers valuable, then, surely, they would help preserve such norms rather than facilitate their abandonment? Although this view seems logical, we will now argue that we should expect norm-compliant robots to encourage potentially unwelcome norm shifts. We will evidence this claim by examining how robots that perform social tasks previously completed exclusively by humans change what it means to do such things.

Let us zoom into norm-compliant robots designed for care. One could imagine that an organization creates a robot that observes the norms nurses generally observe to ensure patients receive proper interpersonal care. Indeed, such machines already exist to some degree (Wright, 2023). This robot respects norms that govern nurses' bedside manner. It behaves as though it understands that it should respect patients' privacy, touch them only when appropriate, and communicate clearly but amicably with them – norms nurses generally respect when interacting with patients (Li, van Wynseberghe, & Roeser, 2020). Even if this

were the case, the robot's presence and behavior will almost certainly encourage its users to develop norms to ensure they can use it.

Patients and human caregivers would have to adapt to this machines' capabilities. Although it may behave like a nurse it cannot do many things that they can and must do. Nurses administer medicine, check patients' vitals, and deal with emergencies. (Contemporary) robots simply do not have the capabilities to attend to such high-risk tasks. Thus, human nurses will continue to perform them. This means that everyone who interacts with this norm-compliant robot must know what it can and cannot do to ensure patients receive proper care. For instance, patients would have to remember that they should not ask this nurse-like robot to increase their dosage of painkillers and, therefore, should instead command it to call a human nurse when they need someone to perform this task or do so themselves. Likewise, nurses would have to learn that they should not leave patients alone with this robot for too long because it cannot attend to their medical needs even though it behaves like someone who can (van Wynsberghe & Li, 2019).

Learning such things would take time and may ultimately result in the emergence of norms that nurses and patients find troubling. Patients may discover that they preferred to communicate exclusively with human nurses if it meant they did not have to constantly bear in mind what they should and should not ask this robot to do. Likewise, nurses may find that reminding forgetful patients who regularly use this robot that they should call them when they need medical attention is more trouble than it is worth and does not save them time at all (van Wynsberghe & Li, 2019). Indeed, many people who use or work alongside this robot may sense that things were better before its introduction.

This discussion draws attention to another issue that deserves recognition. Robots are not humans. Although a nurse-like robot may simulate what it is like to interact with a human nurse with good bedside manner, it cannot be this person and its introduction may undermine norms that govern what care should look like. Although it is norm-compliant, said robot may call into question whether care is something humans should exclusively provide. If we delegated such tasks to robots, we would have to accept the proscription that "it is acceptable to let robots act as care-givers" – a principle that we have good reasons to reject as it may lead to the normalization of care given by unfeeling, unthinking machines.

We contend that robots designed to stand-in for humans in other domains will produce this outcome too. For instance, norm-compliant law-enforcement robots that behave like police officers would problematize norms associated with who (or in this case what) may sanction citizens when they break the law (Calo, 2011). Likewise, norm-compliant robots designed for educational purposes may disrupt what counts as good teaching (e.g., is it something that only qualified, experienced humans should provide? Or something that a machine that can search the internet for answers without knowing what this information means can and should do?) (Sharkey, 2016). Allowing norm-compliant robots to fulfil social roles of this kind may normalize the idea that robots can and should perform such tasks - even though they cannot do many things we expect from the human beings these machines stand in for (e.g., they cannot genuinely care about patients nor understand the value of teaching). Therefore, compelling us to accept norms that prescribe: “one should let a robot do X social task despite its inability to carry out this task as a human would”.

We contend that norm-compliant robots will induce norm changes. Research clearly demonstrates that technological innovation, including robotics, fosters the development and abandonment of norms. These changes often do not amount to a step in the right direction. Indeed, they can undermine norms that we want to preserve. As such, norm-compliant robots can and almost certainly will contribute to the emergence of new norms that may represent a worse way of doing things.

6.4. Conclusion

We aimed to introduce our readers (many of whom we assume work within robot ethics and social robotics) to critical discussions on norm-compliance and demonstrate why we cannot uncritically rely on norms to build ethically-sound robots. We argued that observing a norm does not mean one has acted well. Indeed, in many cases, we have good reasons to claim the opposite. As such, a robot that observes norms may produce outcomes its users, other stakeholders, and society writ large do not endorse. We contend that discussions of this kind need to be more present in the literature on norm-compliant robots and that researchers from this field generally assume that humans and robots should observe norms. We aimed to convince our readers to think otherwise about norms and develop a resource (our “seven troubles with norms”) that other researchers can use to identify potential ethical or political issues raised by norm-compliant robots.

We will now suggest some preliminary mitigation strategies that other researchers could develop to alleviate the issues we identified. First and foremost, we highly recommend that anyone committed to developing norm-compliant robots integrates relevant sociological and political scholarship, some of which we cited in this contribution, into their research. As stated throughout this paper, we did not discover the “seven troubles with norms” we cataloged. Instead, we were the first to apply them to norm-compliant robots. For instance, political theorists have debated how to avoid tyrannies of the majority and paternalism for over two centuries via principles designed to ensure people can enjoy their lives without being unfairly subjugated to other people’s wills and interests. We will not recount these principles here for the sake of brevity. However, we can recommend two contemporary texts that explicitly and implicitly address these issues respectively: namely John Rawls’ *A Theory of Justice* (Rawls,1999)⁹ and Catherine D’Ignazio and Lauren F. Klein’s *Data Feminism* (D’Ignazio & Klein, 2020)¹⁰. Furthermore, we have referenced several useful sociological and philosophical works in this contribution that attempt to develop strategies to help communities abandon flawed norms, most notably Cristina Bicchieri’s *Norms in the Wild* (Bicchieri, 2017) and Geoffrey Brennan, Lina Eriksson, Robert E. Goodin, and Nicholas Southwood’s *Explaining Norms* (Brennan et al., 2013).

Secondly, some of the troubles we identified could be solved or, at the very least, ameliorated via technical means. For instance, one could address some of the challenges raised in the sections on outdated norms and robot-induced norm change by developing robots that update their norm catalogue over time. A possible mitigation strategy for the challenge of outdated norms, and norm change, in general, is to make a robot more sensitive to changes in the social environment. New approaches to norm-aligned robot behavior, like reinforcement learning with normative prior models (Nahian et al., 2021) mentioned in the section on outdated norms, could be adapted to enable the robot to update its norms. One idea here is to continuously train the robot with new material to update the training data with sources that represent the current norms of society.

And lastly; we would like to make clear that many of the issues we have raised cannot be solved solely via technical means. Nonetheless, roboticists could attempt to lessen their effects

⁹ Specifically, in Chapter I: Justice as Fairness and Chapter II: The Principle of Justice.

¹⁰ Specifically, p.21–73.

by incorporating relevant social scientific methods into their research. Adapting participatory and co-design strategies to build norm-compliant robots could help address the challenge of pluralistic ignorance and outdated norms. In participatory and co-design methods, users, and stakeholders are involved in technology design (Steen, 2013; Sanders & Stappers, 2008). Developers of norm-compliant robots could adapt these design approaches so that potential users and other stakeholders give input on the norms a robot is supposed to learn. For instance, robot developers could use focus groups and discussions to investigate people's attitudes about norms (e.g., ask people whether they truly endorse a norm or observe it because they believe their peers expect them to). Researchers could encourage discussions and deliberation about norms to find out which norms are undesired. Additionally, we highly recommend that anyone who wishes to develop norm-compliant robots practice inclusive design (Walsh & Wronsky, 2019; Clarkson et al., 2003) that includes the voices of marginalized populations. Such inclusive co-design strategies would ideally ensure that minority positions about norms are included, which can help to mitigate the issue of tyranny of the majority.

As concluding remark, we would like to add that we do not believe that other researchers should abandon their efforts to develop norm-compliant robots. This is a fascinating and worthwhile endeavor that may well lead to the creation of robots that benefit their users and society writ large. Instead, we aimed to inspire other researcher to think more critically about norms via this contribution to ensure that this comes to pass.

8 Conclusion

8.1. Is hell other people (and now robots)?

I briefly considered calling this thesis *Hell is Other People (and Now Robots)*. Although I ultimately chose another less provocative title, I still believe this allusion to Jean-Paul Sartre's infamous quote (Sartre, 1944) conveys the core message of my PhD research – albeit with more pessimism than I now realize is warranted. Please allow me to explain why. Social interactions are costly. Completing quotidian activities that require coordination efforts tires us out. We rely on highly sophisticated communication skills to meet unstated, communal expectations during many, if not most, of the engagements we enter. I have provided numerous examples of this phenomenon throughout this thesis. Spending all our waking hours attempting to manage how others perceive us would devastate our minds and bodies. In Sartre's words, it would be hell. Privacy protects us from this outcome. We must sporadically remove ourselves from the “whirlwind of daily life” to live well (Westin, 2015, p.41). When alone or among people with whom we share close bonds, we can let our guards down. We do not need to present ourselves as dedicated workers or well-to-do-citizens during these moments. We get to be something other than the impressions we cultivate in public settings.

I have shown that robots, especially those that appear to understand human communication, add new layers of complexity to their users' lives. Robots that simulate what it is like to interact with other people will generate social pressures comparable to those we endure during our engagements with humans. I dedicated much of this thesis to evidencing this hypothesis. For instance, I argued in Chapter 1, that robots installed inside homes will encourage their users to behave as though someone with whom they do not share a private relationship was present. Likewise, I demonstrated in Chapter 2, that robotic “friends” may convince their users to consider entrusting these unthinking machines with information that only someone who cares about them should know. Additionally, I and my co-author Steffen Steinert, contended in Chapter 5, that robots which observe norms will reinforce or reproduce social and political issues associated with norm-compliance that plague our interactions with others despite behaving according to some users' or stakeholders wishes. The robots mentioned above and throughout this thesis represent a new frontier of engagement that threatens to make our lives even more complicated than they already are.

It takes a lifetime of experience to learn how to coordinate with other people successfully and we still make mistakes all the time. Knowing when, where, and with whom we should present ourselves one way rather than another requires us to draw from an ever-growing catalogue of social skills we begin building in childhood and add to indefinitely (Hochschild, 2012, p.3-76). When we meet someone new, we cannot know for certain how our interaction will play out. Although we can infer from context cues what this person expects from us, there is no right way to behave in such situations. Indeed, we must effectively rely on guesswork and hope that this person recognizes our actions' intended meaning, thus, responds to them as we assumed they would. Even if this happens, they may react by saying or doing things that confuse, alarm, or upset us. Thankfully, other humans experience the psychological impacts of interactions too. They also become distressed by unwelcome or prolonged engagements and will attempt to end them at some point. The same cannot be said for robots unless we take proactive steps to prevent this outcome.

Robots that mimic human social behavior encourage people to exert themselves in the manners sketched in the paragraph above. However, they cannot recognize when someone wants to be left alone or feel this way themselves. Once up and running, they ceaselessly manipulate symbols humans use to hold each other's attention. While we know that other people will eventually want to stop communicating with us, we cannot expect this from robots. They cannot simulate this vital aspect of what it is like to interact with another person because they cannot feel anything – let alone the very human urge to end an engagement. Furthermore, other humans appreciate that observing norms, practices, or conventions can distress us and may even sympathize when we disregard them out of frustration. Robots, in contrast, cannot understand why we would act this way or recognize the adverse impact their algorithmically determined behavior has on us. This is why I considered calling this thesis *Hell is Other People (and Now Robots)*. If our employers, corporate actors, and the state decided to install such robots in every locale we enter throughout our daily lives, we would have to constantly endure new, artificial social pressures generated by objects that cannot experience being subject to these demands. I contend that being surround by such machines would be hellish.

Please notice, though, my usage of the conditional tense. *Would*, here, implies a state-of-affairs that has yet come to pass and can only happen if certain conditions are met. I do not believe that robots necessarily produce this outcome. Instead, I think we can address many of the issues I identified in this thesis by changing how we discuss, design, and deploy them. I

abandoned the title mentioned above for this reason. While researching this thesis, I gathered many insights I could not incorporate into its main chapters that I believe may help steer advancements in robotics down a more ethically, socially, and politically sound path. I will dedicate the rest of this conclusion to outlining these recommendations. These suggestions represent research topics I wish I had addressed in this thesis and hope to expand upon later in my career. Therefore, one may read them as my project's limitations too.

8.2. Recommendations

Designing for non-interaction

Humans use various means to signal to one another that they wish to be left alone. I have discussed some of these strategies in this thesis, mainly in Chapter 1. Vigilant readers may have noticed that I primarily referenced passages from Erving Goffman's 1963 work *Behavior in Public Places* to demonstrate this point (Goffman, 1966). Goffman highlights throughout this book that we construct symbolic spaces in public settings that others know they cannot enter without potentially bothering us. We understand that someone reading a newspaper at a train station probably does not want to strike up a conversation. They have silently communicated by placing this object in front of their faces that they would like to wait for their train by themselves. If we ignore this message, we will risk annoying them. According to Goffman, this person has effectively built an ad-hoc zone of privacy. They use this newspaper to shield themselves from unwelcome engagements (Goffman, 1966, p.38-42).

Goffman identifies many other practices that allow us to turn public spaces into quasi-private ones. When we walk through municipal parks, for instance, we may notice that groups of friends sit in circles, thus, ensuring that onlookers see their backs rather than faces. Likewise, someone in a hurry will use their body to cut lines through crowds and stare straight ahead to avoid making eye contact with others (Goffman, 1966, p.151-179). These bodily gestures generate symbolic barriers that separate us from everyone else. We use them to carve out space that people would otherwise assume they can claim. Passersby generally understand that they should not barge through these boundaries as their creators have signaled that they wish to focus their attention on whatever they are doing. Goffman spends much of *Behavior in Public Spaces* explaining how we manage to get things done without being distracted by other people's presence. I believe anyone who wishes to create robots that do not intrude upon their users could learn a lot from this book.

Indeed, I contend that robots should respect the symbolic boundaries outlined above and only invite people to interact with them when they have unmistakably communicated that they wish to do so. For instance, ensuring that robots installed inside stores or offices remain immobile would prevent these machines from disrupting people who want to pass through these locations as quickly as possible (LeCoppe & Rollet, 2020). Likewise, developers could avoid creating robots that butt into conversations or interrupt people's train of thoughts by designing them to stay silent until someone activates them by pressing a clearly marked button that also turns them off. Although I cannot prove it, I sense that many organizations that manufacture robots would find these somewhat obvious suggestions outlandish. Robots tend to have streamlined, futuristic appearances and include as many features as possible that convey they can function without being controlled by a human operator (e.g., mobility). Furthermore, they do not usually have visible on/off switches, I assume, because this would make them appear less autonomous. These are design choices. Developers can build helpful, engaging robots without them. Indeed, doing so would decrease the likelihood that these machines will breach the symbolic boundaries people construct to keep others from disturbing them.

I have long considered developing a taxonomy of behaviors that people use to signal they want to be left alone to help designers create robots that do not initiate unwelcome interactions. Due to time constraints, I have yet to properly research this topic and can only offer the advice outlined above and elsewhere in this thesis. A colleague of mine, Anna Dobrosovestnova, has suggested calling such strategies "designing for non-interaction" and we aim to develop a paper together soon that introduces more nuanced suggestions than I can provide here. Until then, I recommend that researchers working in robotics reflect upon the following question before they decide to add features to a robot that will prompt people to pay attention to it, namely: will this design choice agitate someone who wishes to focus on something else?

Language matters

I began my PhD in January 2019. At this time, researchers already knew that people tend to respond to human-like technologies as though another human were present. Scholars usually call this phenomenon "anthropomorphism". Even though ethicists have warned time and time again that robots which elicit such responses may harm their users in various ways - several

of which I have discussed or identified myself in this thesis – there are substantially more of these technologies deployed today than when I started studying them. Indeed, numerous companies have released large-language models with chatbot-like interfaces within the past year that have convinced many journalists, scholars, and laypeople that they have cognitive abilities comparable to or exceeding those of a human (Wiel, 2023). This is not true. Machine learning programs do not “think”. They perform statistical inferences. Moreover, such false beliefs exemplify and reinforce a way of thinking that, I contend, we should try our utmost to abandon. Namely, that we may use concepts, language and theories devised to describe human experiences to discuss a technology’s operations without committing category mistakes.

I have gestured towards this contention several times in this thesis and avoided anthropomorphizing robots as much as possible. For instance, I implied throughout Chapter 2 that using sociological theory, created to interpret specific human interactions, produces an inaccurate portrayal of human communication strategies and a robot's replication of them. Likewise, I and Steffen Steinert suggested in Chapter 5 that claiming we can build robots that observe norms demonstrates a fundamental misunderstanding of this word's sociological meaning. Even though I have tried to refrain from using language that only makes sense when applied to humans to discuss robots, I am, nonetheless, guilty of anthropomorphizing them. Saying that robots “act”, “encourage”, or “initiates” implies that such machines have intentions when they do not. Unfortunately, the English language compels us to write and talk this way. Describing an object or process that does something will inadvertently assign it (metaphorical) agency.

Although I appreciate that finding language to describe robots in terms that do not anthropomorphize them is difficult, I believe that researchers should work towards this goal. Indeed, I sense that doing so would mitigate many of the ethical issues I have highlighted in this thesis. For instance, I doubt that many users would consider a robot their friend if they properly understood that these machines do not “talk” but instead string together words that other people have indicated they like hearing (Bender et al. 2021). I have occasionally suggested that we could reframe robot’s outputs in a manner that more accurately describes them. Specifically, I have implied or stated several times that whenever a robot says or does something that conveys it has thoughts, feelings, or perceptions it generates fiction. I did not have enough time to properly develop this idea in this thesis but think that it could lead to interesting and fruitful avenues of research.

Fiction affects us. When we read a novel that we find engaging, we become upset and may even cry during passages that describe a beloved character's death or downfall. Likewise, we may feel relief or pride as we watch the protagonist of a video game finally defeat their nemesis. We know that fictional and real-world events are different. Although both generate emotional responses from us, we do not treat the former as though they genuinely happened. Researchers tend to discuss users' perceptions of a robot's simulation of human communication as though these people were witnessing another person experience something. I think this manner of description is leading us astray. If we described human-robot interactions, especially those that simulate social engagements, as fictional representations of real-world situations we may find it easier to talk about these machines without implying they have cognitive faculties they do not have. For instance, we might say that someone who enjoys spending time with a robotic companion has developed an emotional attachment to this machine comparable to those we develop with fictional characters found within books, video games, or movies. This conceptual framing would allow us to discuss the emotional effects such machines have on their users without treating them as something akin to humans. I hope to return to this topic soon and believe literary and media theory could provide us with the concepts and language we need to accurately portray what happens when a robot affects its users.

Privacy and power

In this thesis, I frequently used universalizing language to describe people's experiences of privacy harms. Such phrasing implies that everyone is equally affected by the issues I identified. This is not the case. Although I have occasionally highlighted this fact, I do not believe I have done it justice. Marginalized people tend to endure graver, more frequent attacks on their privacy than members of dominant groups. For instance, many transgender people cannot go about their day without having to enter awkward or hostile social situations we would prefer to avoid. Whereas cisgender people, especially men, may walk through city streets without anyone paying much attention to them, visibly transgender people regularly must deal with lingering stares and, quite often, harassment (Namaste, 2000, p. 135-157; Faye, 2021, p.17-64). In short, cisgender people are more likely to be let alone than transgender people while among strangers. I barely scratched the surface of the power imbalances that determine how much privacy one can expect due to their group affiliations. I consider this the

greatest limitation of my research; especially since the privacy issues I discussed stand to harm some people more than others.

I highly recommend that other researchers bear this in mind if they intend to develop the ideas found in this thesis further. In Chapter 1, for instance, I argued that robots installed inside homes might prompt their users to present themselves as they would during the interactions they complete in public (e.g., at work). Autistic people may find this particularly distressing as they often mask the pain certain social stimuli cause them to ensure that interlocutors do not treat them as abnormal (Hull et al. 2017). Hence, when a robot makes them feel as though another person were present - especially when they expect privacy - they may experience discomfort that neurotypical people do not have to endure (Keyes, 2020). Likewise, I highlighted in Chapter 2 that robots that seem friendly may encourage users to share biographical information with them that they are afraid of telling other people. If a user is from an oppressed group, the outcome of such disclosures may seriously damage their self-esteem. Suppose a young gay man decides to come out to a robot before telling anyone else this information. Considering the biased nature of most machine learning programs' training data (Weidinger et al. 2021; Monae, 2023, p. 61-111; Birhane, Prabhu, Sang, & Boddeti, 2023), this robot quite plausibly could provide a response that implies that being gay is wrong, therefore falsely communicating to this young man that his identity deserves less respect than other people's.

In recent years, numerous researchers from robot ethics and adjacent fields have contended that we must acknowledge the presence of social differences and hierarchies to produce sound ethical research (Costanza-Chock, 2020; Keyes, 2020; D'Ignazio & Klein, 2020, Birhane, 2021). Different groups suffer different harms and injustices mediated via technology. If we fail to appreciate this, we risk erasing entire demographics' experiences of contemporary innovation and its effects on their lives. Aside from ensuring we do not overlook pertinent and pressing ethical issues, integrating diverse perspectives into our research improves its scientific validity, according to many feminist philosophers of science (Haraway, 1998; Harding, 1991; Hill Collins & Bilge, 2016). Knowing how people from politically and academically underrepresented groups perceive phenomenon that we wish to research enables us to develop theory and observations that more adequately represent the objects of our studies (Harding, 1995; Namaste, 2000, p.1-71; D'Ignazio & Klein, 2020, p.149-173). Although more and more researchers working in robot ethics and the philosophy of technology, in general,

accept that we should observe scientific practices that recognize the value of knowledge generated by members of disadvantaged groups, I have encountered very few academic contributions that discuss the effects robots and other adaptive, smart technologies have on marginalized people's sense of privacy. Furthermore, as a field, privacy scholarship rarely discusses oppressed groups' experiences of privacy aside from those attributed to cisgender, heterosexual, usually white women¹¹. I believe my research to date reflects this phenomenological deficit. As such, I soon intend to explore how we can diversify our understanding of privacy to ensure that we can simultaneously identify otherwise inconspicuous privacy problems raised by modern technology and develop a more comprehensive account of this value.

8.3 Concluding remarks

I began this thesis with one primary goal in mind. Namely, to inspire other robot ethicists to think differently about privacy. Exclusively treating this value as the appropriate distribution of information, I contended, would leave us unable to identify many privacy problems raised by robots. Time will tell whether I accomplished this goal. Nonetheless, I hope my peers in robot ethics will see my PhD project as a welcome addition to the field that draws attention to previously under-researched issues we should address if we wish to develop technologies that preserve their users' privacy. On a personal note, my understanding of privacy has radically changed since I started examining this value as a professional academic over four years ago. Learning how scholars have interpreted the meaning of privacy since it became an object of discourse in the mid to late 19th century has significantly altered how I believe we should research this crucial value. Despite being a philosopher by training, I could not have completed this thesis without incorporating sociological, political, historical, and legal insights and theory into my work. I cannot underestimate how rewarding this process has been for me. Indeed, I discovered that my experience of privacy, as an individual, significantly differs from many other peoples' and vice-versa due to my social identity, relationships, and economic situation. This realization, and the interdisciplinary research that led to it, helped me improve how I interpret concepts as a philosopher, in general, enormously. Moreover, I know now that such work is never really done. As many scholars before me have claimed, privacy is "a fraught concept" (Jarvis Thompson, 1975; Solove, 2008) whose definition is and

¹¹ I am referencing what privacy scholars tend to call "the radical feminist critique of privacy" and intersectional scholar's critiques of it (see: MacKinnon, 1989, p. 184-195; Hill Collins, 2000, p.49-76; Ahmed, 2006, p.25-65.)

probably will always remain unsettled. I do not find this well-known observation disheartening. Accepting that we will never fully understand a concept, yet vigorously studying, interpreting, critiquing, and questioning its meaning, nonetheless, I believe, is what philosophers, including robot ethicists, should do. I hope my thesis communicates this ethos and will inspire others to reflect upon what they mean when they use the word “privacy”.

9 Bibliography

- Ackroyd, P. (2017). *Queer City: Gay London from the Romans to the Present Day*. Chatto & Windus.
- Ahmed, S. (2006). *Queer phenomenology: Orientations, objects, others*. Duke University Press.
- Allport, G. (1955). *The Nature of Prejudice*. The Beacon Press.
- Ariès, P. (1962). *Centuries of Childhood*. Trans: Baldick, R. London: Random House.
- Article MATH
- Awad, E., Sohan, D., Kim, R., Schulz, J., Henrich, S., Bonnefon, J., & Rahwan, I. (2018). The Moral Machine Experiment. *Nature* 563, pp. 59-64. <https://doi.org/10.1038/s41586-018-0637-6>
- Babcock, L., Peyser, B., Vesterlind, L., & Wiengart, L. (2022). *The No Club: Putting a stop to women's dead-end work*. London: Simon & Schuster.
- Barcaro, R., Mazzoleni, M. and Virgili, P. (2018). Ethics of care and robot caregivers. *Prolegomena*. 17 (1), pp.71-80. [10.26362/20180204](https://doi.org/10.26362/20180204).
- Bench-Capon, T., & Modgil, S. (2017). Norms and value based reasoning: Justifying compliance and violation. *Artificial Intelligence And Law*, 25, 29–64. <https://doi.org/10.1007/s10506-017-9194-9>.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Bicchieri, C. (2005). *The Grammar of Society: The Nature and Dynamics of Social norms*. Cambridge University Press.
- Bicchieri, C. (2017). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.
- Bicchieri, C., & McNally, p. (2018). Shrieking Sirens—Schemata, Scripts, and social norms: How Change occurs. *Social Philosophy and Policy*, 35(1), 23–53. <https://doi.org/10.1017/S0265052518000079>.
- Birhane, A., Prabhu, V., Han, S., & Boddeti, V. N. (2023). On Hate Scaling Laws For Data-Swamps. <https://doi.org/10.48550/ARXIV.2306.13141>

- Bittman, M., Rice, J. and Wajcman, J. (2004) 'Appliances and their impact: the ownership of domestic technology and time spent on household work', *British Journal of Sociology*, 55, 3, pp. 401–23. <https://doi.org/10.1111/j.1468-4446.2004.00026.x>
- Bourdieu, P. (2010) (tr. Nice, R.) *Distinction: A Social Critique of the Judgement of Taste*, Routledge, Milton Park.
- Bourdieu, p. (2013). *Outline of a theory of practice* (R.Nice , Trans). Cambridge University Press. (Original work published 1977)
- Bouk, D. (2017). The History and Political Economy of Personal Data over the Last Two Centuries in Three Acts. *Osiris*, 32(1), 85–106. <https://doi.org/10.1086/693400>
- Breazeal, C. (2003). Toward sociable robots. *Robotics and Autonomous Systems*, 42(3–4), 167–175. [https://doi.org/10.1016/S0921-8890\(02\)00373-1](https://doi.org/10.1016/S0921-8890(02)00373-1).
- Brennan, G., Eriksson, L., Goodin, R. E., & Southwood, N. (2013). *Explaining norms* (1st ed.). Oxford University Press.
- Brinck, I., Balkenius, C., & Johansson, B. (2016). Making Place for Social Norms in the Design of Human-Robot Interaction. In J. Seibt, M. Nørskov, & S. Schack Anderson (Eds.) *What Social Robots Can and Should Do*. Vol. 290 of *Frontiers in Artificial Intelligence and Applications*, pp. 303–312. IOS Press. <https://doi.org/10.3233/978-1-61499-708-5-303>
- Brown, P. (1991). Passing: Differences in Our Public and Private Self. *Journal of Multicultural Social Work*, 1(2), 33–50. https://doi.org/10.1300/J285v01n02_03
- Bugeja, J., Jacobsson, A., & Davidsson (2021). PRASH: A Framework for Privacy Risk Analysis of Smart Homes. *Sensors (Basel)*; 21(19):6399. doi: 10.3390/s21196399.
- Burns, S. (1977) *The Household Economy: Its Shape, Origins, and Future*, Beacon Press, Boston.
- Burton, N. and Gaskin, J. (2019) ““Thank you, Siri”: politeness and intelligent digital assistants’, *Proceedings of the Americas Conference on Information Systems*, Cancún, Mexico, 15–17 August.
- Butler, J. (2007). *Gender Trouble Feminism and the Subversion of Identity*. (3rd edition). New York: Routledge Classics.
- Calo, R. (2010). People can be so fake: A new dimension to privacy and technology scholarship. *Penn State Law Review*, 114(8), 09–55. <http://papers.ssrn.com/abstract=1458637>.
- Calo, R. (2011) ‘Robots and privacy’ in Lin, P., Abney, K., Bekey, G., Allen, C. and Scheutz, M. (eds) *Robot Ethics*, MIT Press, Cambridge MA, pp. 187–203.

- Čapek, K., & Novack, C. (2004). *R.U.R. (Rossum's universal robots)*. Penguin Books.
- Carlucci, F. M., Nardi, L., Iocchi, L., & Nardi, D. (2015). Explicit representation of social norms for social robots. 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 4191–4196. <https://doi.org/10.1109/IROS.2015.7353970>
- Carmen, A. (2019) 'They welcomed a robot into their family, now they are mourning its death', *The Verge*, available at <https://www.theverge.com/2019/6/19/18682780/jibo-death-server-updates-social-robot-mourning> (accessed March 2022).
- Carros, F., Schwaninger, I., Preussner, A., Randall, D., Wieching, R., Fitzpatrick, G. & Wulf, V. (2022). Care Workers Making Use of Robots: Results of a Three-Month Study on Human-Robot Interaction within a Care Home. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 631, 1–15. <https://doi.org/10.1145/3491102.3517435>
- Carter, E., Reig, S., Tan, X., Laput, G., Rosenthal, S. and Steinfeld, A. (2020) 'Death of a robot: social media reactions and language usage when a robot stops operating', *Proceedings of the 2020 ACM/IEEE International Conference on Human–Robot Interaction*, Cambridge, 23–26 March, pp. 589–97. <https://doi.org/10.1145/3319502.3374794>.
- Castells, M. (1997). *The information age: Economy, society and culture*. Blackwell.
- Caudwell, C. and Lacey, C. (2019) 'What do home robots want? The ambivalent power of cuteness in robotic relationships', *Convergence: The International Journal of Research into New Media Technologies*, 26, 2. [10.1177/1354856519837792](https://doi.org/10.1177/1354856519837792).
- Cerrato, J. & Cifre, E (2018) *Gender Inequality in Household Chores and Work-Family*
- Chatzidakis, A., Hakim, J., Littler, J., Rottenberg, C., and Segal, L. (2020). *The Care Manifesto*. New York: Verso.
- Chen, Y. F., Everett, M., Liu, M., & How, J. p. (2017). Socially Aware Motion Planning with Deep Reinforcement Learning. <https://doi.org/10.48550/ARXIV.1703.08862>
- Clarkson, J., Keates, S., Coleman, R., & Lebbon, C. (Eds.). (2003). *Inclusive Design*. Springer London. <https://doi.org/10.1007/978-1-4471-0001-0>
- Coeckelbergh, M. (2010). Artificial Companions: Empathy and Vulnerability Mirroring in Human-Robot Relations. *Studies in Ethics, Law, and Technology*. 4(3). [10.2202/1941-6008.1126](https://doi.org/10.2202/1941-6008.1126)

- Coeckelbergh, M. (2017). How to describe and evaluate “deception” phenomena: recasting the metaphysics, ethics, and politics of ICTS in terms of magic and performance taking a relations and narrative turn. *Ethics and Information Technology* (2018) 20:71-65 <https://doi.org/10.1007/s10676-017-9441-5>
- Coeckelbergh, M. (2019). *Moved by Machines: Performance Metaphors and the Philosophy of Technology*. Routledge.
- Coggins, T. (2023). Called Back Onstage: Dramaturgic Analysis, Privacy, and Domestic Social Robots. In R. Hakli, p. Mäkelä, & J. Seibt (Ed.), *Social Robots in Social Institutions. Frontiers of Artificial Intelligence and Applications*, pp. 325–334. IOS Press: Amsterdam. <https://doi.org/10.3233/FAIA220632>
- Coggins, T.N. (2022). More work for Roomba? Domestic robots, housework and the production of privacy. *Prometheus*. Vol. 38(1). DOI: 10.13169/prometheus.38.1.0098
- Collingridge, D. (1980). *The Social Control of Technology*. London: Frances Pinter Ltd.
- Costanza-Chock, S. (2020). *Design justice: Community-led practices to build the worlds we need*. The MIT Press.
- Council of Europe (1950) *European Convention for the Protection of Human Rights and Fundamental Freedoms*, 4 November, available at <https://www.refworld.org/docid/3ae6b3b04.html> (accessed August 2021).
- Cox, P. (2013). Passing as Sane, or How to Get People to Sit Next to You on the Bus. In J. A. Brune & D. J. Wilson (eds), *Passing and Disability: Blurring the Lines of Disability* (pp. 99-111) Temple University Press.
- Craig L, Mullan K (2011) How Mothers and Fathers Share Childcare: A Cross-National Time-Use Comparison. *American Sociological Review* 76(6): 834–861.
- Crawford, K. (2021) *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, Yale University Press, New Haven.
- Crawford, K. and Joler, V. (2018) *Anatomy of an AI System: The Amazon Echo as an Anatomical Map of Human Labor*, AI Now Institute and Share Laboratory, available at <https://anatomyof.ai> (accessed March 2022).
- Crossley, S. (2017). *In Their Place: The Imagined Geographies of Poverty*. London: Pluto Press.
- Cuijpers, C & Koops, B. (2013). Smart Metering and Privacy in Europe: Lessons from the Dutch Case. In: S. Gutwirth et al. (eds), *European Data Protection: Coming of Age*,

- Dordrecht: Springer, pp. 269-293 (2012), Available at SSRN: <https://ssrn.com/abstract=2218553>
- D'Ignazio, C. & Klein, L.F. (2020) *Data Feminism*. MIT Press: Cambridge
- Danaher, J. (2017). The symbolic-consequences argument in the sex robot debate. In J. Danaher, & N. McArthur (Eds.), *Robot sex: Social and ethical implications*. Cambridge: MIT Press.
- Danaher, J. (2019). The philosophical case for robot friendship. *Journal of Posthuman Studies*, 3(1), 5–24.
- Darling, K. (2012). Extending Legal Rights to Social Robots. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2044797>
- Dautenhahn, K. (2007). Socially intelligent robots: Dimensions of human–robot interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480), 679–704. <https://doi.org/10.1098/rstb.2006.2004>.
- Davis, A.Y. (2019). *Women, Race and Class*. Penguin Books. London.
- De Tocqueville, A. (2010). *Democracy in America*. (J.T. Schleifer, Trans). Liberty Fund. (Original work published 1835).
- De Wildt, T. E., Van De Poel, I. R., & Chappin, E. J. L. (2022). Tracing Long-term Value Change in (Energy) Technologies: Opportunities of Probabilistic Topic Models Using Large Data Sets. *Science, Technology, & Human Values*, 47(3), 429–458. <https://doi.org/10.1177/01622439211054439>
- DeGraaf, M. (2016). An Ethical Evaluation of Human–Robot Relationships. *Int J of Soc Robotics* (2016) 8:589–598. 10.1007/s12369-016-0368-5
- Dickinson, E. (1970). *The complete poems of Emily Dickinson* (T. H. Johnson, Ed.). Faber and Faber.
- Dobrosovetsnova, A., & Hannibal, G. (2021). Working alongside service robots: challenges to workplace identity performance. *Culturally sustainable social robotics*. In: *Proceedings of Robophilosophy 2020*, vol 335, pp 148–157
- Dobrosovetsnova, A., Hannibal, G. and Reinboth, T. (2021) ‘Service robots for affective labor: a sociology of labor perspective’, *AI & Society*, April, available at <https://doi.org/10.1007/s00146-021-01208-x> (accessed March 2022).
- Dorlin, E. (2022). *Self-Defense: A Philosophy of Violence*. (K. Aarrons, Trans). Verso. London. (Original Work published 2019).
- Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, 42(3–4), 177–190. [https://doi.org/10.1016/S0921-8890\(02\)00374-3](https://doi.org/10.1016/S0921-8890(02)00374-3).

- Dworkin, G. (2005). Moral Paternalism. *Law and Philosophy*, 24(3), 305–319. <https://doi.org/10.1007/s10982-004-3580-7>.
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109(3), 573–598. <https://doi.org/10.1037/0033-295X.109.3.573>.
- Elder, A. (2015). False Friends and False Coinage: A tool for navigating the ethics of sociable robots. *SIGCAS Comput. Soc.* 45, 3. <https://doi.org/10.1145/2874239.2874274>
- Elster, J. (2014). Tyranny and brutality of the majority. In J. Elster, & S. Novak (Eds.), *Majority decisions. Principles and Practices* (pp. 159–176). Cambridge University Press.
- Engels, F. (2010). *The Origins of the Family, Private Property and the State*. (A. West, Trans.) London: Penguin Books Ltd. (Original work published 1884).
- Etherington, D. (2016) ‘iRobot says 20 percent of the world’s vacuums are now robots’, *Techcrunch.com*, available at <https://techcrunch.com/2016/11/07/irobot-says-20-percent-of-the-worlds-vacuums-are-now-robots/> (accessed August 2021).
- European Court of Human Rights. (2022). *Guide on Article 8 of the European Convention on Human Rights - Right to respect for private and family life, home and correspondence.* Council of Europe. https://www.echr.coe.int/documents/guide_art_8_eng.pdf.
- Faye, S. (2021). *The Transgender Issue: An Argument for Justice*. Milton Keynes: Penguin Random House UK.
- Federici, S. (2012) *Revolution at Point Zero: Housework, Reproduction, and Feminist Struggle*, PM Press, Oakland CA.
- Feinberg, J. (1989). *The moral limits of the criminal law. 3: Harm to self*. Oxford Univ. Pr.
- Firestone, S. (2015). *The Dialectics of Sex: The Case for Feminist Revolution*. London: Verso.
- Forlizzi, J. (2008). The product ecology: understanding social product use and supporting design culture. *International Journal of Design*, 2, 1, pp.11–20.
- Forlizzi, J., & DiSalvo, C. (2006). Service robots in the domestic environment: A study of the roomba vacuum in the home. *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, 258–265. <https://doi.org/10.1145/1121241.1121286>
- Fortunati, L. (2018). Robotization and the domestic sphere. *New Media & Society*, 20(8), 2673–2690. <https://doi.org/10.1177/1461444817729366>
- Foucault, M (1975). *Discipline and Punish*. (A. Sheridan, Trans). Vintage Books: New York.

- Fraser, N. (1987). Women, Welfare and the Politics of Need Interpretation. *Hypatia* 2(1), pp.103-121.
- Fraser, N. (1994). After the Family Wage: Gender Equity and the Welfare State. *Political Theory* 22(4), pp.591-618.
- Friedman, B., Kahn, P., Hagman, J., Severson, R., & Gill, B. (2006). The Watcher and the Watched: Social Judgments About Privacy in a Public Place. *Human-Computer Interaction*, 21(2), 235–272. https://doi.org/10.1207/s15327051hci2102_3
- Fuse, Y., Takenouchi, H., & Tokumaru, M. (2019). A Robot Model That Obeys a Norm of a Human Group by Participating in the Group and Interacting with Its Members. *IEICE Transactions on Information and Systems*. E102.D. 185-194. [10.1587/transinf.2018EDP7077](https://doi.org/10.1587/transinf.2018EDP7077).
- Garreau, J. (2007). Bots on the Ground. *Washington Post*. https://www.washingtonpost.com/wp-dyn/content/article/2007/05/05/AR2007050501009_pf.html (Retrieved 15/09/2023)
- Garbes, A. (2022). *Essential Labor: Mothering as Social Change*. New York: Harper Collins.
- Garfinkel, H. (1967). *Studies in Ethnomethodology*. Prentice-Hall
- Gertz, N. (2018). Hegel, the Struggle for Recognition, and Robots. *Techne: Research in Philosophy and Technology* 22(2) pp.138-157. <https://doi.org/10.5840/techne201832080>
- Glancy, D.J. (1979). The Invention of the Right to Privacy. *Arizona Law Review* 21(1) pp.1-39.
- Gobetti, D. (1997) *Humankind as a System: Private and Public Agency at the Origins of Modern Liberalism*. In (eds) Weintraub, J and Kumar, K. *Public and Private in Thought and Practice: Perspectives on a Grand Dichotomy*. Chicago: The University of Chicago Press, p.103-133
- Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Penguin Books; 1959.
- Goffman, E. (1963). *Stigma: Notes on the Management of Spoiled Identity*. Penguin Books.
- Goffman, E. (1966). *Behaviour in Public Places: Notes on the Social Organisation of Gatherings*. The Free Press. New York.
- Graham, S., & Thrift, N. (2007). Out of Order: Understanding Repair and Maintenance. *Theory, Culture and Society*, 24(3). <https://doi.org/10.1177/0263276407075954>
- Grandey, A.A. (2003). When “the show must go on”: surface acting and deep acting as determinants of emotional exhaustion and peer-rated service delivery. *Academy of Management Journal*, 46(1), 86–96. <https://doi.org/10.2307/30040678>.

- Granfield, R. (1991). Making it by Faking it: Working-Class Students in an Elite Academic Environment. *Journal of Contemporary Ethnography*, 20(3), 331–351. <https://doi.org/10.1177/089124191020003005>
- Gray, M. and Suri, S. (2019) *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*, Mariner Books, Boston.
- Grill, K., & Hanna, J. (Eds.). (2018). *The Routledge handbook of the philosophy of paternalism*. Routledge.
- Guizzo, E. (2016) ‘The little robot that could . . . Maybe: Jibo is as good as social robots get. But is that good enough?’, *IEEE Spectrum*, 53, 1, pp.58–62. Doi:10.1109/MSPEC.2016.7367471.
- Habermas, J. (1992). *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society* (T. Burger & F. Lawrence, Trans.). Polity Press. London. (Original work published 1962)
- Hamed, A.S., Lalanne, D., and Rogers, Y. (2020) . The Five Strands of Living Lab: A Literature Study of the Evolution of Living Lab Concepts in HCI. *ACM Trans. Comput.-Hum. Interact.* 27, 2, pp.1-26 <https://doi.org/10.1145/3380958>
- Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, 14(3), 575. <https://doi.org/10.2307/3178066>
- Hayes-Roth, B. (2004). What Makes Characters Seem Life-Like. In H. Prendinger & M. Ishizuka (Eds.), *Life-Like Characters* (pp. 447–462). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-08373-4_19
- Heidegger, M. (1996). *Being and Time*. (J. Stambaugh, Trans.). State University of New York Press. New York. (Original work published 1953)
- Hentschel, T., Braun, S., Peus, C., & Frey, D. (2018). The communality-bonus effect for male transformational leaders – leadership style, gender, and promotability. *European Journal of Work and Organizational Psychology*, 27(1), 112–125. <https://doi.org/10.1080/1359432X.2017.1402759>
- Herman, J.L. (2015). *Trauma & Recovery: The Aftermath of Violence – From Domestic Abuse to Political Terror*. Basic Books. New York.
- Hill Collins, P. (2000). *Black Feminist Thought*. Routledge. New York City.
- Hill Collins, P., & Bilge, S. (2016). *Intersectionality*. Polity Press.
- Hochschild, A.R. (1979). Emotion Work, Feeling Rules, and Social Structure. *American Journal of Sociology* 85(3): pp.551-575.

- Hochschild, A. (2012). *The Managed Heart: Commercialization of Human feeling* (3rd ed.). London: University of California Press.
- Hochschild, A., & Machung, A. (1989). *The second shift: working parents and the revolution at home*. New York City: Penguin Group.
- Honneth, A. (2004). *The Struggle for Recognition: The Moral Grammar of Social Conflicts*. Trans: Anderson, J. Cambridge: Polity Press.
- Horne, C., & Mollborn, S. (2020a). Norms: An Integrated Framework. *Annual Review of Sociology*, 46(1), 467–487. <https://doi.org/10.1146/annurev-soc-121919-054658>.
- Kolodny, L. (January 2017). Hanson Robotics Built a Professor Einstein Toy to Teach Kids Science with a Familiar Face. *Techrunch.com*. <https://techcrunch.com/2017/01/23/hanson-robotics-built-a-professor-einstein-toy-to-teach-kids-science-with-a-familiar-face/>
- Hustinx, P. (2010). Privacy by design: Delivering the promises. *Identity in the Information Society*, 3(2), 253–255. <https://doi.org/10.1007/s12394-010-0061-z>
- Inness, J. (1992) *Privacy, Intimacy, and Isolation*, Oxford University Press, Oxford.
- International Federation of Robotics (2019) ‘Executive summary world robotics 2019: service robots’, available at https://ifr.org/downloads/press2018/Executive_Summary_WR_Service_Robots_2019.pdf (accessed April 2021).
- Intille, S., Larson, K., Munguia Tapia, E., Beaudin, J.S., Kaushik, P., Nawyn, J. and Rockinson, R. (2006). Using a Live-In Laboratory for Ubiquitous Computing Research. In: Fishkin, K.P., Schiele, B., Nixon, P., Quigley, A. (eds) *Pervasive Computing*. Pervasive 2006. Lecture Notes in Computer Science. 3968, pp.349- 365 https://doi.org/10.1007/11748625_22
- Jackson, R. B., & Williams, T. (2019). Language-Capable Robots may Inadvertently Weaken Human Moral Norms. 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 401–410. <https://doi.org/10.1109/HRI.2019.8673123>
- Jibo (2014) ‘Jibo, the world’s first social robot for the home’, available at <https://www.indiegogo.com/projects/jibo-the-world-s-first-social-robot-for-the-home#/> (accessed October 2021).
- Johnson, A. G. (2014). *The gender knot: Unraveling our patriarchal legacy* (Third Edition). Temple University Press.
- Kahn, P. H., Ishiguro, H., Friedman, B., Kanda, T., Freier, N. G., Severson, R. L., & Miller, J. (2007). What is a Human?: Toward psychological benchmarks in the field of

- human–robot interaction. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, 8(3), 363–390. <https://doi.org/10.1075/is.8.3.04kah>
- Kanuha, V.K. (1999). The Social Process of “Passing” to Manage Stigma: Acts of Internalised Oppression or Acts of Resistance? *Journal of Sociology and Social Welfare*, 6(4).
- Katz, D., & Allport, F. H. (1931). *Students’ attitudes; a report of the Syracuse University reaction study*. Craftsman Press.
- Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: State of the art, current trends and challenges. *Multimedia Tools and Applications*, 82(3), 3713–3744. <https://doi.org/10.1007/s11042-022-13428-4>
- Kolodny, L. (January 2017). *Hanson Robotics Built a Professor Einstein Toy to Teach Kids Science with a Familiar Face*. Techcrunch.com.
- Koops, B. J., Newell, B., Timan, T., Skorvánek, I., Chokrevski, T., & Galič, M. (2017). A typology of privacy. *University of Pennsylvania Journal of International Law*, 38(2), 483-575. <http://ssrn.com/abstract=2754043>
- Kudina, O. (2021). “Alexa, who am I?”: Voice Assistants and Hermeneutic Lemniscate as the technologically mediated sense-making. *Human Studies*, 44(2), 233–253. <https://doi.org/10.1007/s10746-021-09572-9>.
- Kudina, O. and Coeckelbergh, M. (2021), ““Alexa, define empowerment”: voice assistants at home, appropriation and technoperformances”, *Journal of Information, Communication and Ethics in Society*, 19(2), pp. 299-312. <https://doi.org/10.1108/JICES-06-2020-0072>
- Kudina, O., & Verbeek, P.-P. (2019). Ethics from Within: Google Glass, the Collingridge Dilemma, and the Mediated Value of Privacy. *Science, Technology, & Human Values*, 44(2), 291–314. <https://doi.org/10.1177/0162243918793711>
- Kumar, K. (1997) ‘Home: the promise and predicament of private life at the end of the twentieth century’ in Weintraub, J. and Kumar, K. (eds) *Public and Private in Thought and Practice: Perspectives on a Grand Dichotomy*, University of Chicago Press, Chicago, pp. 204–37.
- Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers Through Society*. Harvard University Press. Cambridge, Massachusetts
- Leenes, R., Palmerini, E., Koops, B.-J., Bertolini, A., Salvini, P., & Lucivero, F. (2017). *Regulatory challenges of robotics: Some guidelines for addressing legal and ethical*

- issues. *Law, Innovation and Technology*, 9(1), 1–44.
<https://doi.org/10.1080/17579961.2017.1304921>
- Lerner, G. (1987) *The Creation of Patriarch*, Oxford University Press, Oxford.
- Levin, I. (1972) *The Stepford Wives*, Corsair, London.
- Li, S. (2022). Reconceptualizing Autonomy in Elderly Care in the Robot Era: A Relational Perspective. <https://doi.org/10.4233/uuid:91c17930-79ad-4c6f-921f-3870c2f1a33d>
- Li, S., van Wynsberghe, A., & Roeser, S. (2020). The Complexity of Autonomy: A Consideration of the Impacts of Care Robots on the Autonomy of Elderly Care Receivers. In M. Nørskov, J. Seibt, & O. S. Quick (Eds.), *Culturally Sustainable Social Robotics—Proceedings of Robophilosophy 2020* (pp. 316–325). IOS Press. <https://doi.org/10.3233/FAIA200928>
- Licoppe, C., & Rollet, N. (2020). « Je dois y aller » (pp. 220–221). *Réseaux*. 2–3 <https://doi.org/10.3917/res.220.0151>. Analyses de séquences de clôtures entre humains et robot.
- Lin, Y. et al. (2021) ‘The mental health of transgender and gender non-conforming people in China: a systematic review’, *The Lancet Public Health*, 6(12), pp. e954–e969. Available at: [https://doi.org/10.1016/S2468-2667\(21\)00236-X](https://doi.org/10.1016/S2468-2667(21)00236-X).
- Linnet, T. (2020). Exploitation as innovation: research ethics and the governance of experimentation in the urban living lab. *Regional Studies*. 55 (12) pp.1902-1912. <https://doi.org/10.1080/00343404.2020.1826421>
- Lutz, C., Schöttler, M., & Hoffmann, C. P. (2019). The privacy implications of social robots: Scoping review and expert interviews. *Mobile Media & Communication*, 7(3), 412–434. <https://doi.org/10.1177/2050157919843961>
- Mackinnon, C. (1991). *Toward a Feminist Theory of State*. London: Harvard University Press.
- Malle, B. F. (2016). Integrating robot ethics and machine morality: The study and design of moral competence in robots. *Ethics and Information Technology*, 18(4), 243–256. <https://doi.org/10.1007/s10676-015-9367-8>.
- Malle, B. F., & Scheutz, M. (2014). Moral competence in social robots. 2014 IEEE International Symposium on Ethics in Science, Technology and Engineering, 1–6. <https://doi.org/10.1109/ETHICS.2014.6893446>
- Malle, B. F., Rosen, E., Chi, V. B., Berg, M., & Haas, p. (2020). General Methodology for Teaching Norms to Social Robots. *The 29th IEEE International Conference on Robot and Human Interactive Communication*.

- Marion Young, I. (2005). *On Female Body Experience: “Throwing Like a Girl” and Other Essays*. New York City: Oxford University Press.
- Mattern, S. (2019, October 2). *Minimal Maintenance*. Lapsus Lima. <https://www.lapsuslima.com/minimal-maintenance/>
- McMunn, A., Bird, L., Webb, E., and Sacker, A. (2019). Gender Divisions of Paid and Unpaid Work in Contemporary UK Couples. *Work, Employment and Society* 34, pp.155 - 173.
- Mill, J. S. (1985). *On liberty*. Penguin Classics. (Originally published 1859)
- Miller, D. T., & McFarland, C. (1987). Pluralistic ignorance: When similarity is interpreted as dissimilarity. *Journal of Personality and Social Psychology*, 53(2), 298–305. <https://doi.org/10.1037/0022-3514.53.2.298>.
- Mokrosinska, D. (2018). Privacy and Autonomy: On Some Misconceptions Concerning the Political Dimensions of Privacy. *Law and Philosophy*, 37(2), 117–143. <https://doi.org/10.1007/s10982-017-9307-3>
- Monea, A. (2023). *The digital closet: How the internet became straight*. MIT Press.
- Muishout, C., Coggins, T. and Schipper, H. (2020) ‘More than meets the eye? Robotisation and normativity in the Dutch construction industry’, *Proceedings of the Second RILEM International Conference on Concrete and Digital Fabrication*, July, pp. 839–51. https://doi.org/10.1007/978-3-030-49916-7_82.
- Nahian, M. S. A., Frazier, S., Harrison, B., & Riedl, M. (2021). Training Value-Aligned Reinforcement Learning Agents Using a Normative Prior. <https://doi.org/10.48550/ARXIV.2104.09469>
- Nahian, M. S. A., Frazier, S., Riedl, M., Harrison, B. Learning Norms from Stories: A Prior for Value Aligned Agents. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (2020). 124–130. <https://doi.org/10.1145/3375627.3375825>
- Namaste, V. K. (2000). *Invisible lives: The erasure of transsexual and transgendered people*. University of Chicago Press.
- Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, 56(1), 81–103. <https://doi.org/10.1111/0022-4537.00153>
- Nyholm, S. (2020) *Humans and Robots: Ethics, Agency, and Anthropomorphism*, Rowman & Littlefield, Lanham MD.
- Nyirkos, T. (2020). *The tyranny of the majority: History, concepts, and challenges*. Routledge.
- O’Gorman, H. J. (1979). White and black perceptions of racial values. *Public Opinion Quarterly*, 43(1), 48. <https://doi.org/10.1086/268490>.

- O’Gorman, H. J. (1986). The discovery of pluralistic ignorance: An ironic lesson. *Journal of the History of the Behavioral Sciences*, 22(4), 333–347.
- O’Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Penguin Books.
- Okita, S. and Ng-Thow-Hing, V. (2014) The effects of design choices on human–robot interactions in children and adults. in Markowitz, J. (ed.) *Robots that Talk and Listen: Technology and Social Impact*, De Gruyter, Berlin.
- Oldenziel, R. (2001). Man the Maker, Woman the Consumer: The Consumption Junction Revisited. In A.N.H. Creager, E. Lunbeck & Londa Schiebinger, (Eds.) *Feminism in Twentieth Century Science, Technology, and Medicine* (pp.128-148). Chicago: Chicago University Press.
- Oliver, M. (1990). *The Politics of Disablement*. Palgrave Macmillan: New York.
- Pasquale, F. (2020). *New laws of robotics: Defending human expertise in the age of AI*. The Belknap Press of Harvard University Press.
- Penney, J. (2021). Understanding Chilling Effects. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3855619>
- Penz, O. & Sauer, B. (2019) *Governing affects: neoliberalism, neo-bureaucracies, and service work*. Routledge.
- Perkins-Gillman, C. (2020). *The Home, its Work and Influence*. Alpha Editions.
- Prentice, D. A., & Miller, D. T. (1996). Pluralistic Ignorance and the Perpetuation of Social Norms by Unwitting Actors. In *Advances in Experimental Social Psychology* (Vol. 28, pp. 161–209). Elsevier. [https://doi.org/10.1016/S0065-2601\(08\)60238-5](https://doi.org/10.1016/S0065-2601(08)60238-5)
- Prost, A. (1998). Public and Private Spheres in France. In (eds) Ariès, P. and Duby, G., *A History of Private Life, vol. 5: Riddles of Identity in Modern Times*, Harvard University Press, Cambridge MA, pp.1-103
- Prost, A. (1998) ‘Public and private spheres in France’ in Ariès, P. and Duby, G., *A History of Private Life, vol. 5: Riddles of Identity in Modern Times*, Harvard University Press, Cambridge MA.
- Puig de la Bellacasa, M. (2017). *Matters of Care: Speculative Ethics in More than Human Worlds*. University of Minneapolis Press. London.
- Rawls, J. (1999). *A Theory of Justice*, revised edition. The Belknap Press of Harvard University Press. Cambridge, Massachusetts.
- Real, T. (1997). *I Don’t Want to Talk About It: Overcoming the Secret Legacy of Male Depression*. Scribner. New York.

- Renfrow, D. (2011). A Cartography of Passing in Everyday Life. *Symbolic Interaction*, 27(4), 485–506. <https://doi.org/10.1525/si.2004.27.4.485>
- Riaz, F., Jabbar, S., Sajid, M., Ahmad, M., Naseer, K., & Ali, N. (2018). A collision avoidance scheme for autonomous vehicles inspired by human social norms. *Computers & Electrical Engineering*, 69, 690–704. <https://doi.org/10.1016/j.compeleceng.2018.02.011>.
- Rogers, M. (1992). They Were All Passing: Agnes, Garfinkel, and Company. *Gender & Society*, 6(2), 169–191. <https://doi.org/10.1177/089124392006002002>
- Rössler, B. (2007). *The Value of Privacy*. Trans: Glasgow, R.D.V. Cambridge: Polity. (Original work published 2001)
- Rössler, B. (2007). Work, Recognition, Emancipation. In (eds) den Brink, B. and Owen, D. *Recognition and Power: Axel Honneth and the Tradition of Critical Social Theory*. Cambridge: Cambridge University Press, pp.135-164.
- Russel, A.L., & Vinsel, L. (2016). Hail the Maintainers. *Aeon.com*. <https://aeon.co/essays/innovation-is-overvalued-maintenance-often-matters-more>
- Ryland, H. (2021). It's Friendship, Jim, but Not as We Know It: A Degrees-of-Friendship View of Human–Robot Friendships. *Minds and Machines*, 31(3), 377–393. <https://doi.org/10.1007/s11023-021-09560-z>
- Sanders, E. B. N., & Stappers, p. J. (2008). Co-creation and the new landscapes of design. *CoDesign*, 4(1), 5–18. <https://doi.org/10.1080/15710880701875068>.
- Sartre, J.-P. (1989). *No exit, and three other plays* (Vintage International ed). Vintage International.
- Scheuerman, M. K., Hanna, A., & Denton, E. (2021). Do Datasets Have Politics? Disciplinary Values in Computer Vision Dataset Development. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 1–37. <https://doi.org/10.1145/3476058>
- Scheutz, M. (2009) 'The inherent dangers of unidirectional emotional bonds between humans and social robots' in Lin, P., Abney, K., Bekey, G., Allen, C. and Scheutz, M. (eds) *Robot Ethics*, MIT Press, Cambridge MA, pp.205–23.
- Schwab, K. (2017) *The Fourth Industrial Revolution*, Penguin, Harmondsworth.
- Schwartz Cowan, R. (1985). *More Work for Mother: The Ironies of Household Technology from the Open Hearth to the Microwave*. Basic Books.
- Sharkey, A.J.C & Sharkey, N. (2020). We need to talk about deception in social robotics! *Ethics and Information Technologies*, 23 (pp. 309-316).

- Sharkey, A.J.C. (2016). Should we welcome robot teachers?. *Ethics Inf Technol* 18, 283–297 (2016). <https://doi.org/10.1007/s10676-016-9387-z>
- Sharkey, N & Sharkey, A. (2010). The crying shame of robot nannies An ethical appraisal. *Interaction Studies - INTERACT STUD.* 11. 10.1075/is.11.2.01sha.
- Shibata, T., & Wada, K. (2011). Robot Therapy: A New Approach for Mental Healthcare of the Elderly – A Mini-Review. *Gerontology*, 57(4), 378–386. <https://doi.org/10.1159/000319015>
- Siebers, T. (2004). Disability as Masquerade. *Literature and Medicine*, 23(1), 1–22. <https://doi.org/10.1353/lm.2004.0010>
- Smith, M. (2019, October 5). Most smart speaker owners are rude to their devices. YouGov. <https://yougov.co.uk/topics/technology/articles-reports/2019/05/10/most-smart-speaker-owners-are-rude-their-devices>
- Solove, D. J. (2006). A Taxonomy of Privacy. *University of Pennsylvania Law Review*, 154(3), 477. <https://doi.org/10.2307/40041279>
- Solove, D. (2008) *Understanding Privacy*, Harvard University Press, Cambridge MA.
- Solove, D. (2012). Privacy as Self-Management and the Consent Dilemma. *Harvard Law Review*, 126.
- Song, V. (2017) ‘Jibo review’, *PC Mag*, available at <https://www.pcmag.com/reviews/jibo> (accessed October 2021).
- Sparrow, R. & Sparrow, L. (2006). In the hands of machines? The future of aged care. *Minds and Machines* 16, 141-161.
- Steen, M. (2013). Co-Design as a process of Joint Inquiry and Imagination. *Design Issues*, 29(2), 16–28. https://doi.org/10.1162/DESI_a_00207.
- Stilgoe, J. (2017). Seeing Like a Tesla: How Can We Anticipate Self-Driving Worlds? *Glocalism*, 3, DOI: 10.12893/gjcp.2017.3.2
- Stone, S. (1987). *The Empire Strikes Back: A Posttranssexual Manifesto*. In J. Epstein & C. Straub (eds), *Body Guards: the Cultural Politics of Gender Ambiguity* (pp.280-304). New York: Routeldge
- Strasser, S. (1982) *Never Done: A History of American Housework*, Holt Paperbacks, New York.
- Stryker, S. (2017). *Transgender History: The Roots of Today’s Revolution* (2nd Edition). Seal Press.
- Sung, J.-Y., Guo, L., Grinter, R. and Christensen, H. (2007) “‘My Roomba is Rambo’”: intimate home appliances’ in Krumm, J., Abowd, G., Seneviratne, A. and Strang, T.

- (eds) UbiComp 2007, Springer-Verlag, Berlin, pp. 145–62.
https://doi.org/10.1007/978-3-540-74853-3_9.
- Sunstein, R. (1996). Social Norms and Social Roles. *Columbia Law*, 96(4) pp. 903-968.
- Swierstra, T. (2015). Identifying the normative challenges posed by technology’s ‘soft’ impacts. *Etikk i Praksis - Nordic Journal of Applied Ethics*, 9(1),
<https://doi.org/10.5324/eip.v9i1.1838>.
- Swierstra, T., Stemerding, D., & Boenink, M. (2009). Exploring Techno-Moral Change: The Case of the Obesity Pill. In p. Sollie & M. Düwell (Eds.), *Evaluating New Technologies. Methodological Problems for the Ethical Assessment of Technology Developments*, (pp. 119–138). Springer. https://doi.org/10.1007/978-90-481-2229-5_9
- Terpstra, A., Schouten, A. P., De Rooij, A., & Leenes, R. E. (2019). Improving privacy choice through design: How designing for reflection could support privacy self-management. *First Monday*. <https://doi.org/10.5210/fm.v24i7.9358>
- Thomson, J. J. (1975). The Right to Privacy. *Philosophy & Public Affairs*, 4(4), 295–314.
<http://www.jstor.org/stable/2265075>
- Thorp, J. (1992). The Social Construction of Homosexuality. *Phoenix*, 46(1), 54–61.
<https://doi.org/10.2307/1088774>
- Tomic, S., Pecora, F., & Saffiotti, A. (2018). Norms, Institutions, and Robots. *ArXiv:1807.11456 [Cs]*. <http://arxiv.org/abs/1807.11456>
- Tordoff, D.M., Wanta, J.W., Collin A, Stepney, C., Inwards-Breland, D.J., and Ahrens, K. (2022). Mental Health Outcomes in Transgender and Nonbinary Youths Receiving Gender-Affirming Care. *JAMA Netw Open*. 2022;5(2):e220978.
doi:10.1001/jamanetworkopen.2022.0978
- Tronto, J. (1993). *Moral Boundaries: A Political Argument for an Ethics of Care*. Routledge. London.
- Tronto, J. (2013). *Caring Democracy: Markets, Equality, and Justice*. New York City: New York University Press.
- Turban, J. L. and Ehrensaft, D. Research Review: Gender identity in youth: treatment paradigms and controversies. (2017) *J Child Psychol and Psychiatr*.
doi:10.1111/jcpp.12833.
- Turkle, S. (2011) *Alone Together: Why we Expect More from Technology and Less from Each Other*, Basic Books, New York.

- Turkle, S. (2017) 'A nascent robotics culture: new complicities for companionship' in Wallach, W. and P. Asaro, P. (eds) *Machine Ethics and Robot Ethics*, Routledge, Milton Park, pp. 107–16. <https://doi.org/10.4324/9781003074991-12>.
- Turkle, S., Taggart, W., Kidd, C. and Dasté, O. (2006) 'Relational artifacts with children and elders: the complexities of cybercompanionship', *Connection Science*, 18, 4, pp. 347–61. <https://doi.org/10.1080/09540090600868912>
- Turow, J. (2017). *The aisles have eyes: How retailers track your shopping, strip your privacy, and define your power*. Yale University Press.
- Ulanoff, L. (2017) 'Jibo is a cute home robot, but it's not as smart as Alexa', Mashable, available at <https://mashable.com/article/jibo-review> (accessed October 2021).
- Ureta, J., Brito, C., Dy, J., Santos, K., Villaluna, W. and Ong, E. (2020) 'At home with Alexa: a tale of two conversational agents' in Sojka, P., Kopecek, I., Pala, K. and Horak, A. (eds) *Proceedings of the 23rd International Conference on Text, Speech, and Dialogue*, Brno, 8–11 September, Springer-Verlag, Berlin, pp. 495–503.
- Urquhart, L., Reedman-Flint, D., & Leesakul, N. (2019). Responsible domestic robotics: Exploring ethical implications of robots in the home. *Journal of Information, Communication and Ethics in Society*, JICES-12-2018-0096. <https://doi.org/10.1108/JICES-12-2018-0096>
- Urry, J. (2005). *The Tourist Gaze* (2nd edition). Sage Publications.
- Van Camp, J. (2017) 'Review: Jibo social robot', *Wired*, available at <https://www.wired.com/2017/11/review-jibo-social-robot/> (accessed October 2021).
- van Camp, J. (2019) 'My Jibo is dying and it is breaking my heart', *Wired*, available at <https://www.wired.com/story/jibo-is-dying-eulogy/> (accessed October 2021).
- van de Poel, I. (2016). An ethical Framework for evaluating Experimental Technology. *Science and Engineering Ethics*, 22(3), 667–686. <https://doi.org/10.1007/s11948-015-9724-3>.
- van De Poel, I. (2021). Design for value change. *Ethics and Information Technology*, 23(1), 27–31. <https://doi.org/10.1007/s10676-018-9461-9>
- van Den Hoven, M. J. (1997). Privacy and the varieties of moral wrong-doing in an information age. *ACM SIGCAS Computers and Society*, 27(3), 33–37. <https://doi.org/10.1145/270858.270868>
- van Wynsberghe, A. and Li, S. (2019a) 'A paradigm shift for robot ethics: from HRI to human–robot–system interaction (HRSI)', *Medicolegal and Bioethics*, 9, pp. 11–21. <https://doi.org/10.2147/MB.S160348>.

- van Wynsberghe, A., & Li, S. (2019). A paradigm shift for robot ethics: from HRI to human–robot–system interaction. *Medicolegal and Bioethics*:11–21.
- Vanek, J. (1974) ‘Time spent in housework’, *Scientific American*, 231, 5, pp. 116–21, available at <http://www.jstor.org/stable/24950221> (accessed August 2021).
- Vaussard, F., Fink, J., Bauwens, V., Rétornaz, P., Hamel, D., Dillenbourg, P. and Mondada, F. (2014) ‘Lessons learned from robotic vacuum cleaners entering the home ecosystem’, *Robotics and Autonomous Systems*, 62, 3, pp. 376–91. <https://doi.org/10.1016/j.robot.2013.09.014>.
- Vinsel, L., & Russel, A.L. (2020). *The Innovation Delusion: How our Obsession with the New has Disrupted the Work that Matters Most*. Penguin Random House. New York.
- Waldman, A. E. (2018). *Privacy as trust: Information privacy for an information age*. Cambridge University Press.
- Wallach, W., & Allen, C. (2008). *Moral machines: Teaching robots right from wrong*. Oxford University Press.
- Walsh, G., & Wronsky, E. (2019). AI + Co-Design: Developing a Novel Computer-supported Approach to Inclusive Design. Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing, 408–412. <https://doi.org/10.1145/3311957.3359456>
- Wang, X., Shen, J., and Chen, Q. (2021). How PARO can help older people in elderly care facilities: A systematic review of RCT. *International journal of nursing knowledge*. 3310.1111/2047-3095.12327.
- Warren, S and Brandeis, L. (1890) The right to Privacy. *Harvard Law Review*, 4, 5, pp.193–220.
- Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L. A., ... Gabriel, I. (2021). Ethical and social risks of harm from Language Models. <https://doi.org/10.48550/ARXIV.2112.04359>
- Weil, E. (2023) You Are Not a Parrot: And a chatbot is not a human. And a linguist named Emily M. Bender is very worried what will happen when we forget this. *New York Magazine*. <https://nymag.com/intelligencer/article/ai-artificial-intelligence-chatbots-emily-m-bender.html> (Retrieved 15/09/2023).
- Westin, A. F. (2015). *Privacy and freedom* (New edition). IG Publishing.

- Wiederhold, B. (2018) “‘Alexa, are you my mom?’ The role of artificial intelligence in child development”, *Cyberpsychology, Behavior and Social Networking*, 21, 8, pp. 471–2. <https://doi.org/10.1089/cyber.2018.29120.bkw>.
- Wright, J. (2023). *Robots won’t save Japan: An ethnography of Edlercare automation*. Cornell University Press, ILR Press.
- Young, M. (2021). Maintainence. In (eds) Michelfelder, D and Doorn, N. *The Routledge Handbook of the Philosophy of Engineering*. New York City: Taylor & Francis, p.356-369
- Zheng, W., Surgevil, O., & Kark, R. (2018). Dancing on the Razor’s edge: How top-level women leaders manage the paradoxical tensions between Agency and Communion. *Sex Roles*, 79(11–12), 633–650. <https://doi.org/10.1007/s11199-018-0908-6>.
- Zuboff, S. (2019). *Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books Ltd.

10 Summary

In the introduction of this thesis, I contend that robot ethics, as a research field, generally treats privacy as the appropriate distribution of information, and therefore overlooks privacy concerns raised by robots beyond this conceptualization's purview. I illustrate this contention by evaluating a hypothetical case involving a household companionship robot via contemporary robot ethics literature focusing on privacy. I argue that this corpus cannot identify a variety of privacy concerns raised by such robots because it relies on a narrow interpretation of privacy that can only recognize privacy harms of an informational nature. I posit that privacy represents considerably more than implied by the interpretations offered by robot ethicists. Most crucially, it signifies our need to withdraw sporadically from social engagements. Considering that robots - like the one described in the case mentioned - simulate what it is like to interact with other humans, I argue that such machines will produce privacy concerns when they successfully create the impression that another person is present during moments when their users wish to be left alone. I highlight that some researchers from robot ethics have discussed issues of this kind but rarely frame them as privacy concerns, thus leaving a significant literature gap I attempt to fill via my research. I conclude the introduction by presenting a close reading of relevant privacy scholarship to evidence the claims made above and lay the theoretical foundation for the dissertation.

In Chapter 1, I argue that social robots installed inside homes produce a novel privacy problem when they invite their users to engage with them. I synthesize two bodies of literature to accomplish this goal, namely: Erving Goffman's early sociological work on character performances, and privacy scholarship which argues that humans must withdraw from their public commitments to live well. Goffman and the privacy scholars I cite contend that humans spend much of their everyday lives attempting to meet social expectations they do not control. Goffman calls such strategies "character performances" and clarifies that being among others who expect us to behave in such ways compels us to do so (Goffman, 1959, p.13-15). Furthermore, he argues that we may conceptualize the spaces we enter throughout our day-to-day lives as either "onstage" or "offstage" locales (Goffman, 1959, p. 109-141). The former represents places where we anticipate we must remain in character, while the latter symbolizes environments where we cease such effort to attend to tasks that we cannot adequately complete during performances. I draw from various privacy scholars' work to frame our homes as "offstage" environments which should enable us to attend to needs we cannot satisfy

in public, onstage settings, which I label as self-care, autonomy, and intimacy. I argue that robots which elicit social responses from their users prompt them to perform in character as they would during interactions with the people they encounter in public, onstage setting, thus robbing them of time they could dedicate to the needs mentioned above – a phenomena I categorize as a privacy problem.

In Chapter 2, I identify privacy problems robots produce when they simulate what it is like to interact with a friend. I begin by reconstructing an argument I call “the performance account” that other philosophers developed to advocate for the possibility of human-robot friendships. The performance account contends that Erving Goffman, and the sociologists he influenced, would support the idea that we classify someone as our friend when they consistently perform actions that communicate that this is the case. Therefore, we can call a robot our friend if it behaves as we expect a friend to behave. I argue that this hypothesis does not hold water when we consult the sociological literature the performance account uses and that it excuses two privacy problems that friend-like robots create. I interpret the sociological literature on “passing” to evidence these claims. I explain that people who practice passing conceal biographical facts during specific social interactions. Crucially for my argument, they do not always do this and usually tell people they wish to become friends with biographical information they conceal while passing. I introduce a thought experiment involving passing to demonstrate that we must tell someone something true about biographies to become their friend. As robots cannot experience anything, they cannot complete this action; thus, cannot be our friends. I conclude the chapter by demonstrating that friendly seeming robots will mediate surveillance by encouraging users to disclose sensitive biographical information during interactions and rob users of the care and compassion they deserve when they share truths about themselves. I claim that both issues are privacy problems.

In Chapter 3 I examine the labor needed to produce spaces one may call “private” and robots’ impact on these practices. I argue that we should expect robots designed to lessen the amount of housework homemakers complete to change how they perform such labor rather than reduce the time they spend on it. I appeal to numerous feminist scholars' work to demonstrate that we cannot enjoy the privacy afforded by our homes unless someone has produced the conditions necessary to do so, usually via unpaid housework. Women have historically fulfilled this role and labored inside homes without financial compensation, thus leaving them with far less time to dedicate to their privacy than men. Over the past two centuries, many

companies have created technologies that supposedly reduce the need for certain types of housework. Feminist historians posit that these technologies' widespread deployment has increased rather than decreased homemakers' workloads. I identify the key processes that lead to this outcome by interpreting several feminist historians' work. I use these historical insights to evaluate a selection of contemporary robots. I argue that these machines will contribute to the processes mentioned above, thus creating new housework, and changing what their users must do before they can enjoy their private lives.

In Chapter 4, I and my co-author, Madelaine Ley, reflect upon a normative assumption I leave mostly unquestioned throughout this thesis. Namely, that we deserve private spaces, most notably, our homes. We introduce three philosophical accounts of the home and show how ethicists of technology may use them to evaluate domestic innovations' moral import, including robots. We begin by discussing a conceptualization of the home derived from classical liberal theory, we label "the relief account", which treats the home as a sanctuary cut off from the rest of the world, thus enabling occupants to recover the energy they exert in public. Afterwards, we outline another conceptualization of the home, we label "the production account", based upon the work of several Marxist-feminist scholars. This account claims we should treat our homes as economic entities comparable to offices or factories as they provide occupants and, by extension, capitalist economies, with necessary resources. We assess both accounts' validity and highlight their deficiencies. We develop a third conceptualization, inspired by Axel Honneth and Iris Marion Young's work, we label the maintenance account, to address the other two's flaws. We argue that our home lives should help us maintain self-esteem and therefore know that other people and institutions should treat us with respect. We identify two factors that contribute to this outcome, namely, knowing that we share our homes exclusively with people who love us unconditionally and have a space for ourselves that reflects our identity and histories. We conclude the chapter by showing how other ethicists of technology may use the maintenance account to evaluate domestic innovations.

In the final chapter of the thesis, I and my co-author Steffen Steinert, critique an assumption many robot ethicists and researchers from human-robot interaction studies have expressed. Namely, that humans and robots which observe norms produce ethically sound social interactions. We developed this critique because many authors from the fields mentioned above have called for the development of norm-compliant robots without properly examining

how this could go wrong. We begin our critique by providing a sociological definition of norms. We then interpret relevant sociological and political literature to outline what we call “seven troubles with norms” which we argue will derail efforts to build ethically sound norm compliant robots. We label these seven troubles, norm biases, paternalism, tyranny of the majority, pluralistic ignorance, norm abandonment and robot induced norm change. Overall, we claim that developing robots which observe norms will reinforce or reproduce social hardships and ethical issues that many people already face. We conclude by recommending that researchers who wish to develop norm-compliant robots accept that doing so does not necessitate that these machines will seamlessly integrate into our social world and that we should develop means to ensure they do not make everyday interactions any more morally dubious than they already are.

In the thesis’ conclusion, I reflect on how life among other people already causes us unavoidable stress and warn that deploying seemingly socially aware technologies, e.g., robots, in public and private locales will aggravate these tensions. I provide a selection of recommendations, based on my research, to address this issue and hopefully lessen the likelihood that robots will further complicate how we coordinate with one another. I conclude the thesis by meditating on what it means to study privacy and claim that philosophers who wish to understand this value should accept that its definition will always remain unsettled. This conceptual ambiguity, I claim, should inspire us to develop interdisciplinary interpretations of privacy that help us appreciate what this infamously nebulous value represents.

11 Samenvatting

In de inleiding van dit proefschrift stel ik dat het onderzoeksveld van de robotethiek privacy in het algemeen beschouwt als een gepaste verdeling van informatie, en zo geen oog heeft voor privacy-problematiek rondom robots die niet onder deze conceptualisering valt. Ik illustreer deze stelling met een hypothetische casus waarin ik een robot die thuis gezelschap houdt, beoordeel aan de hand van recente literatuur op het gebied van robotethiek gericht op privacy. Ik betoog dat deze werken blind zijn voor verschillende privacyproblemen die bij dergelijke robots voorkomen, omdat ze gebruikmaken van een zeer nauwe interpretatie van privacy die alleen rekening houdt met privacyschending op het gebied van informatie. Maar privacy houdt aanzienlijk meer in dan deze lezingen van robotethici suggereren. Bovenal staat privacy voor onze behoefte om ons zo nu en dan te onttrekken aan sociale verplichtingen. Omdat robots – zoals degene die ik in de casus beschrijf – menselijke interactie simuleren, stel ik dat zulke machines privacy-issues gaan veroorzaken wanneer ze erin slagen de indruk wekken dat er iemand aanwezig is op momenten dat de gebruiker liever alleen is. Ik benadruk dat sommige robotethiek-onderzoekers dit soort kwesties wel hebben besproken, maar zelden onder de noemer ‘privacykwesties’. Hierdoor is de literatuur onvolledig en dat gat probeert mijn onderzoek op te vullen. Ik sluit de inleiding af met een ‘close reading’ van relevante privacy-literatuur om de hierboven gemaakte beweringen te bewijzen en zo de theoretische basis voor het proefschrift te leggen.

In hoofdstuk 1 beargumenteer ik dat sociale robots die in huizen worden geïnstalleerd een nieuw privacyprobleem veroorzaken zodra ze contact zoeken met gebruikers. Hiervoor combineer ik twee verzamelingen literatuur, te weten Erving Goffman’s vroege sociologische werk over het opvoeren van rollen en privacy-literatuur die betoogt dat mensen zich moeten kunnen terugtrekken van hun publieke verplichtingen om gelukkig te zijn. Goffman en de geciteerde privacywetenschappers beweren dat mensen een groot deel van hun dagelijks leven proberen te voldoen aan sociale verwachtingen waar ze geen controle over hebben. Goffman noemt dergelijke strategieën ‘character performances’ (het spelen van rollen) en stelt dat we dit doen omdat anderen van ons verwachten dat we ons zo gedragen (Goffman, 1959, p.13-15). Bovendien stelt hij dat we de ruimtes die we in ons dagelijks leven betreden, kunnen opdelen in ‘op het podium’ of ‘backstage’ (Goffman, 1959, p. 109-141). Onder de eerste categorie vallen ruimtes waarin we verwachten dat we in onze rol moeten blijven; de tweede categorie is voor ruimtes waar we dat niet hoeven om dingen te doen die ons ‘in onze rol’ niet

zouden lukken. Ik put uit het werk van verschillende privacywetenschappers om onze huizen te framen als ‘backstage’-omgevingen die ons zouden moeten helpen aan behoeften te voldoen die we niet in het openbaar ‘op het podium’ kunnen bevredigen en die ik indeel in categorieën als zelfzorg, autonomie en intimiteit. Ik beargumenteer dat robots die aansturen op sociale interactie, gebruikers ertoe aanzetten om dezelfde rol te spelen als tijdens persoonlijke interactie in het openbaar, en ze dus ‘op het podium’ plaatsen, wat tijd wegneemt om aan de hierboven genoemde behoeften te besteden – een fenomeen dat ik als privacyprobleem beschouw.

Hoofdstuk 2 gaat over privacyproblemen die robots veroorzaken als ze vriendschappelijke omgang simuleren. Ik begin met een betoog dat ik ‘de performance-theorie’ noem en door andere filosofen is ontwikkeld om te pleiten voor de mogelijkheid van mens-robotvriendschappen. De performance-theorie stelt dat Erving Goffman, en de sociologen die hij beïnvloedde, het idee onderschrijven dat we iemand als vriend zien als diegene consequent acties uitvoert die duidelijk maken dat het om vriendschap gaat. Daarom kunnen we een robot onze vriend noemen als die zich gedraagt zoals we verwachten dat een vriend zich gedraagt. In mijn optiek houdt deze hypothese geen stand als we de sociologische literatuur met betrekking tot de performance-theorie raadplegen, omdat deze twee privacyproblemen negeert die worden veroorzaakt door vriend-achtige robots. Ik gebruik de sociologische literatuur over ‘passing’ om deze beweringen te staven. Bij ‘passing’ (letterlijk: ervoor doorgaan) verbergen mensen persoonlijke informatie tijdens specifieke sociale interacties. Cruciaal voor mijn betoog is dat ze dit niet voor altijd blijven doen en deze informatie meestal wel delen wanneer ze vrienden met iemand willen worden. Een gedachte-experiment met ‘passing’ toont aan dat we mensen iemand iets echt uit ons persoonlijk leven moeten vertellen om vrienden te worden. Omdat robots niets kunnen ervaren, kunnen ze dit niet doen; en kunnen ze dus ook geen vriendschap sluiten. Ik sluit het hoofdstuk af door aan te tonen dat vriendelijk overkomende robots eigenlijk surveillanten zijn, die gebruikers aanmoedigen gevoelige persoonlijke informatie te delen zonder de zorg en compassie te verlenen die mensen verdienen bij het doen van zulke ontboezemingen. Wat mij betreft zijn beide van deze kwesties privacyproblemen.

In hoofdstuk 3 onderzoek ik de arbeid die nodig is om zogenaamde ‘privé’-ruimtes te scheppen en de invloed die robots hierop hebben. Ik betoog dat robots die zijn ontworpen om huishoudelijke taken te verlichten dat werk veranderen en niet verminderen. Ik haal hierbij

diverse feministische auteurs aan om aan te tonen dat we pas thuis van privacy kunnen genieten als iemand de juiste voorwaarden daarvoor heeft geschapen, meestal door middel van onbetaald huishoudelijk werk. Vrouwen hebben door de geschiedenis heen deze taak op zich genomen, waardoor ze veel minder tijd privé overhielden dan mannen. De afgelopen twee eeuwen hebben veel bedrijven technologieën ontwikkeld om dat werk uit handen te nemen. Feministische historici stellen dat de alomtegenwoordigheid van deze technologieën de werklust eerder heeft vergroot dan verkleind. Ik onderscheid de belangrijkste processen die hiertoe leiden aan de hand van het werk van verschillende feministische historici. Deze geschiedkundige inzichten gebruik ik om een aantal hedendaagse robots te evalueren. Ik beweer dat deze machines ook zo'n effect zullen hebben, waardoor nieuwe huishoudelijke taken ontstaan en invloed hebben op wat gebruikers moeten doen voordat ze van hun privéleven kunnen genieten.

In hoofdstuk 4 reflecteer ik met co-auteur, Madelaine Ley, op een normatieve aanname die ik in dit proefschrift grotendeels ongemoeid laat. Namelijk dat we privéruimtes verdienen, en dan vooral thuis. We introduceren drie filosofische verhandelingen over 'thuis' en laten zien hoe technologie-ethici deze kunnen gebruiken om de morele implicaties van huishoudelijke innovaties, waaronder robots, te evalueren. We beginnen met een conceptualisering van 'thuis', afgeleid van de klassieke liberale theorie. Deze 'opladen-theorie' behandelt het huis als een toevluchtsoord afgesloten van de rest van de wereld, waar bewoners de energie kunnen herwinnen die ze buitenshuis spenderen. Daarna schetsen we een andere conceptualisering van 'thuis' die we 'de productie-theorie' noemen, geïnspireerd op verschillende marxistisch-feministische auteurs. Deze theorie stelt dat we onze huizen moeten zien als economische entiteiten, vergelijkbaar zijn met kantoren of fabrieken, omdat ze bewoners, en in het verlengde daarvan kapitalistische economieën, voorzien van benodigde middelen. We toetsen beide theorieën en benadrukken hun tekortkomingen. Daarnaast ontwikkelen we een derde conceptualisatie, geïnspireerd op het werk van Axel Honneth en Iris Marion Young: de 'onderhoudstheorie', die de minpunten van de andere twee theorieën toont. Wij beweren dat ons thuisleven ons zelfvertrouwen op peil houdt en ons eraan herinnert dat andere mensen en organisaties ons met respect moeten behandelen. We onderscheiden twee factoren die hieraan bijdragen, namelijk het idee dat we onze huizen uitsluitend delen met mensen die onvoorwaardelijk van ons houden en dat het huis als ruimte een afspiegeling is van onze identiteit en geschiedenis. We sluiten het hoofdstuk af door te demonstreren hoe andere

technologie-ethici de ‘onderhoudstheorie’ kunnen gebruiken om huishoudelijke innovaties te evalueren.

In het laatste hoofdstuk van het proefschrift bekritisieren ik met co-auteur Steffen Steinert een veelvoorkomende aanname onder robotethici en onderzoekers naar mens-robotinteractie. Namelijk dat mensen en robots die normen naleven ethisch verantwoorde sociale interacties hebben. Deze kritiek is een reactie op de binnen deze vakgebieden veelgehoorde oproep voor de ontwikkeling van normconforme robots, zonder gedegen onderzoek naar hoe dit fout kan gaan. Deze kritiek start vanuit een sociologische definitie van normen. Vervolgens schetsen we aan de hand van relevante sociologische en politicologische literatuur ‘zeven problemen met normen’, die de bouw van ethisch verantwoorde, normconforme robots in de weg staan. Deze zeven problemen zijn normbias, paternalisme, tirannie van de meerderheid, pluralistische onwetendheid, normvervaging en door robots geïnitieerde normverandering. Over het geheel genomen stellen we dat de ontwikkeling van robots die zich aan normen houden de reeds in het leven aanwezige sociale beproevingen en ethische kwesties zal versterken of reproduceren. We sluiten af met de aanbeveling dat onderzoekers die normconforme robots willen ontwikkelen, moeten accepteren dat deze machines niet noodzakelijkerwijs naadloos in onze maatschappij zullen integreren en dat we er zorg voor moeten dragen dat ze doorsnee interacties niet moreel dubieuzer maken dan dat zijn ze al.

In de conclusie van het proefschrift reflecteer ik op hoe het leven onder de mensen ons al onvermijdelijke stress bezorgt en waarschuw ik dat het inzetten van ogenschijnlijk sociaal vaardige technologieën, zoals robots, deze spanning zal verergeren in zowel het publiek als het private domein. Ik geef een aantal aanbevelingen voor dit probleem, gebaseerd op mijn onderzoek, en hoop zo de kans te verkleinen dat robots de manier waarop we met elkaar omgaan verder zullen bemoeilijken. Ik sluit het proefschrift af door na te denken over wat het betekent om onderzoek te doen naar privacy en stel dat filosofen die deze waarde willen begrijpen, moeten accepteren dat de definitie ervan altijd ongewis zal blijven. Wat mij betreft zou deze conceptuele dubbelzinnigheid ons moeten inspireren om een interdisciplinair begrip van privacy te ontwikkelen dat recht doet aan wat deze notoir vage term allemaal behelst.

11 Acknowledgements

Without the enduring support of the following friends, family members, and colleagues, I would have not completed this dissertation. May life be as kind to you as you have been to me; Menah Wellen, Marie van Loon, Jack Coggins, Madelaine Ley, Steffen Steniert, Shuhong Li, Olya Kudina, Ibo van de Poel, Pim Haselager, Giulio Mecacci, Wouter Mulders, Dylan Cawthorne, Hugo Emmerzael, Sacha Gertsik, Rijk Kistemaker, Leah Heaton-Jones, Mark Thomas Young, Bas de Boer.

12 About the Author

Thomasin Coggins (1990) was born in Newport, the Isle of Wight, the United Kingdom. As of March 2024, she is a Postdoctoral Researcher at Donders Institute for Brain, Cognition, and Behaviour, Radboud University. She completed her PhD in Ethics of Technology at Delft University of Technology between January 2019 and March 2024. She also holds a Bachelor of Arts and Master of Arts (research) in philosophy from Anglia Ruskin University and the University of Amsterdam, respectively. She is also the managing editor of the Journal of Human-Technical Relations and the co-founder of Intersectional Philosophy of Technology research group.

13 List of Publications

- Coggins, T.N. (forthcoming). Friends are not “electric” (characters): a sociological case against human-robot friendships. *Journal of Human-Technology Relations*
- Coggins, T. N., & Steinert, S. (2023). The seven troubles with norm-compliant robots. *Ethics and Information Technology*, 25(2), 29. <https://doi.org/10.1007/s10676-023-09701-1>
- Coggins, T. (2023). Called Back Onstage: Dramaturgic Analysis, Privacy, and Domestic Social Robots. In R. Hakli, p. Mäkelä, & J. Seibt (Ed.), *Social Robots in Social Institutions. Frontiers of Artificial Intelligence and Applications*, pp. 325–334. IOS Press: Amsterdam. <https://doi.org/10.3233/FAIA220632>
- Coggins, T. N. (2022). More work for Roomba? Domestic robots, housework and the production of privacy. *Prometheus*, 38(1). <https://doi.org/10.13169/prometheus.38.1.0098>
- Muishout, C. E., Coggins, T. N., & Schipper, H. R. (2020). More Than Meets the Eye? Robotisation and Normativity in the Dutch Construction Industry. In *Second RILEM International Conference on Concrete and Digital Fabrication* (pp. 839-851). (RILEM Bookseries; Vol. 28). Springer. https://doi.org/10.1007/978-3-030-49916-7_82