



In Love with a Corporation without Knowing It: An Asymmetrical Relationship

Claire Boine, Céline Castets-Renard, Aurélie Clodic, Rachid Alami

► To cite this version:

Claire Boine, Céline Castets-Renard, Aurélie Clodic, Rachid Alami. In Love with a Corporation without Knowing It: An Asymmetrical Relationship. Nørskov, M., Seibt J., Quick O. 2020. Culturally Sustainable Social Robotics-Proceedings of Robophilosophy 2020., IOS Press, 2020, Frontiers of AI and Its Applications,, 10.3233/FAIA200923 . hal-02963070

HAL Id: hal-02963070

<https://hal.science/hal-02963070>

Submitted on 26 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

In love with a corporation without knowing it: an asymmetrical relationship

Claire BOINE ^{a,1}, Céline CASTETS-RENARD ^b, Aurélie CLODIC ^c and Rachid ALAMI ^c

^a*Boston University, School of Public Health / University of Ottawa, Faculty of Law*

^b*University of Ottawa, Faculty of Law / Artificial and Natural Intelligence Toulouse Institute*

^c*LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France / Artificial and Natural Intelligence Toulouse Institute*

Abstract. Most areas of the law rely on the assumption on free will, which manifests in the expression of consent. We examine the nature of human emotions toward fictional characters and social robots, and question the concept of consent in the context of these unreciprocated fictional relationships. We conclude that policies need to regulate the use of social robots in order to protect consumers, and especially vulnerable ones, from an asymmetry of power between them and robotic companies. We propose different statutory and design-based solutions depending on the purpose of the robots and the type of users.

Keywords. Social robots, sex robots, emotions, human robot interaction, consumer protection, ethics, robot design, free will, consent

1. Introduction

Most areas of the law rely on the assumption on free will. In criminal law, sentences are most often justified by retributive justice, the idea that criminals must proportionately suffer in return, and by deterrence theory, which states the prospect of being punished deters individuals from committing crimes in the first place. Both theories assume that criminals could have chosen to not commit their crime but did not. Without free will, retributive justice would simply be cruel and deterrence ineffective. In contract law, contracts are valid if they are not signed under coercion. Outside of the protection of vulnerable persons, the law assumes that individuals make decisions freely on a daily basis. Economic liberalism also relies on free will because the market allocation relies on agents being intrinsically rational and allocating value to goods based on what they are willing to pay. Nonetheless, legislators have also recognized that there is a power differential between individuals and companies, and that consumers cannot exercise their free will if they do not have access to complete and fair information about products. Consumer law is an attempt to compensate for this imbalance. For instance, cigarette manufacturers must label the packages with health warnings to make sure consumers are aware of the adverse risks of the product for their health [1].

However, a new technology recently introduced on the market is creating an unprecedented power asymmetry between companies and individuals: social robots.

¹ Community Health Sciences, Boston University School of Public Health, 801 Massachusetts avenue, Boston MA 02118, United States; e-mail: cboine@bu.edu

Although definitions of social robots vary, most agree that they are robots that interact with humans in socially acceptable ways. So far, social robots have been introduced as companions in nursing homes, hospitals, offices, and private homes around the world. Some of them purely look like objects (e.g. Jibo), some have animal-like appearances (e.g. Paro the seal, Robear...) and others are anthropomorphic (e.g. Buddy, Lynx, Mark I, Sofia, Erica...) [2, 3, 4, 5, 6, 7, 8]. Still far from achieving general artificial intelligence or having sentience, social robots can work in various capacities such as administrative assistants, home security systems, social companions, elderly care workers—Robear lifts patients in Japanese nursing homes—and child sitters. Several have also been introduced as sex robots (e.g. Samantha) or are being developed for that purpose [9]. Due to social robots being quite expensive, their use is not yet widespread. We can expect them to become more affordable as technology improves in the next few years.

The proliferation of social robots in the near future poses multiple ethical challenges. Sparrow has argued that the use of social robot for companionship is morally wrong because it relies on sentimentalism, i.e. choosing to self-delude into thinking that robots are sentient [10]. He argues that pet robot owners choose to wrongly believe that their robots have affection for them. To him, this behavior violates a rule according to which one ought to apprehend the world accurately. Rodogno, who disagrees, responds that in most cases, social robot users do not engage in sentimentality [11]. They are not deluded, he writes, and simply choose to suspend their beliefs in the same way as readers do with novels. In addition, even if users of social robots purposefully chose to mischaracterize the world, Rodogno argues that such sentimentality would not in itself cause harm to anyone. The robot is not harmed by their owner purposefully ignoring that they are not real, and neither is the owner. However, the author is concerned with the harms that the use of social robots on a large scale could cause. Just like one car was harmless but the mass production of cars contributed to climate change, he believes that the widespread use of robot companions could adversely transform human relationship. He especially worries that socialization with robots may render humans incapable of negotiating alterity and loving others with their imperfections.

This paper will not address the topic of morality. However, we will build on Rodogno's work because the question of whether social robot users engage in sentimentality is of paramount importance to the legal field. His paper especially summarizes a long-standing debate in the humanities: can fiction actually make us feel authentic emotions? And if so, what do they rely on? We will draw from the fields of neuroscience and evolutionary psychology to show that the way our brain has evolved makes us vulnerable when using social robots. In particular, the unreciprocated emotional attachment that can be rooted in human-robot interaction creates an asymmetrical power dynamic between individuals and companies. We will then look into the ethical questions that arise from the power differential, both in consumers' relationships to their social robots and to others. Then, using American and European precedents, we will turn to the law and how it has dealt with similar challenges in the past to suggest ways law makers can protect individuals from the potential harms of social robots.

2. The risk of social robots in light of evolutionary psychology

2.1. Emotions and fiction

Both 18th century rationalism and the growing importance of evolutionary biology in the 19th century left us with the misconceived idea that emotions would be a vestige of our pre-historic brain and would not be fit for the modern world. To emotions, perceived as irrational, we oppose reason and self-restrain. Evolutionary psychologist Steven Pinker, noting that our genes have undergone strong selection in the past few millennia, has contended that “the mind is adapted to a mixture of recent and ancient environments” [12. p. xii]. Regretting the use of the term “reptilian brain” when it comes to our emotions, Pinker demonstrates that emotions are “mechanisms that set the brain’s highest-level goals” [12 p. 373]. They intervene in the same time as other components of our thought process and prompt us to act. Furthermore, although there is a misconception that the selfish gene theory implies that humans only care about access to food, personal security and sexual reproduction, Pinker insists that understanding the environment and securing the cooperation of others are high on the list of human priorities. In short, emotions are a tool for our genome to motivate us to act in ways that will help us obtain resources, such as social cooperation, that are necessary to its preservation.

The humanities have long tried to understand the mystery of readers’ emotional involvement in fictional stories. The fact that we could actually feel sad or happy for characters who do not exist in real life puzzled many. Some have argued that the emotions we feel for fictional characters are not real, while others have posited that when we feel empathy toward a character, it is actually directed at real people who would be in the same situation. Rodogno contends that the actual existence of a character is not a necessary condition for emotions to arise [11]. After all, we would not enjoy watching a movie or reading a novel if we did not care about the content on the basis that none of it is true. We would probably read encyclopedias and history textbooks to children before bedtime instead of tales.

However, recent findings in evolutionary psychology and neuroscience have shed light on the critical role of fiction for human development. Harari shows that our capacity to believe in fiction is what distinguishes us from other animals, and what put us at the top of the food chain [13]. Fiction, he posits, allow us to cooperate with complete strangers around common purposes. Our beliefs in companies, nations, religions, and common values have enabled us to build social structures that would have been impossible otherwise. Harari argues that from the very beginning, our ability to gossip and tell stories promoted our cooperation. Humans are the most social animals, and their survival rests upon the group and social cooperation. We might be the most gullible of all animals, accepting to believe white lies on a daily basis in order to maintain social relations.

On top of that, we now know that fiction plays a major role in human development of social skills and empathy. The development of first-person and epistolary novels would have made readers more empathetic [14,15]. This hypothesis has been tested in several experiments. Mar et al. showed that exposure to fiction predicts the score on a well-accepted measure of empathy—being able to recognize others’ emotions by only seeing their eye region on pictures [16]. Djikic et al. conducted a study in which half of the participants had to read *The Lady With the Toy Dog* by Chekhov, while the other half read a comparison text that contained the same content in the form of an administrative report [17]. The treatment group had experienced emotions while

reading, which led them to being more open to novelty, whereas the control group reading the report had not.

More importantly, brain imaging has brought a start of explanation to this relationship. Speer et al. exposed study participants to fiction while in an MRI [18]. Their experiment shows that when the participants hear about a character performing an action, the regions of their brains that activate are the same as if they performed the action themselves. Our brains might simulate the stories we hear so that we can better make sense of what others go through and relate to them, in order to promote cooperation. This is in line with Pinker's predicted human goals to understand the environment and secure the cooperation of others [12]. In other words, Sapiens having empathy and being able to both share stories and relate to others' may have created more frequent or meaningful connections that may have led to their survival. Regardless of the evolutionary reason behind that fact, these recent experiments show that the participants experience the same emotions when hearing the story and when performing the action themselves. Knowing that a story is fictional does not preclude them from experiencing real emotions. These findings thus render the question of belief irrelevant and discredit the theory of quasi-emotions.

These scientific discoveries are critical to the understanding of our emotional life in relation to non-living artifacts. As Jerome Bruner wrote, "the distinction between narrative fiction and narrative truth is nowhere nearly as obvious as common sense and usage would have us believe" [19]. It is important to remember the tenuous boundary between fiction and reality when we consider our relationship to social robots.

2.2. Fake friends and real love

If the most recent developments in our genome are from the past few millennia, needless to say we have not yet adapted to interacting with anthropomorphic robots. Readers might counter-argue that individuals are not only a biological product, and that they are rational agents with intentions of their own. Whether or not this is the case is certainly not relevant here. We merely argue that our emotional responses are bound by some biological constraints that arose in an environment in which only living creatures were animated. Since social robots were introduced, we have heard of several instances in which people got emotionally attached to them. Most famously, iRobot, the company that produces the autonomous vacuum cleaners called Rumba, was surprised to realize that their customers would not let them replace their defective devices. Their customers had become emotionally attached to these specific units, and thus projected individuality and personality onto them [20]. After getting attached to their vacuum cleaner, some owners had even changed their initial beliefs and were convinced that their device had specific unique personality traits. In an experiment, Darling et al. witnessed scientists refuse to destroy robots that looked like baby dinosaurs that were crying [21]. Members of the military have also grown attached to combat robots. An officer stopped a demining mission because he could not stand seeing the demining robot lose any more leg [21]. In some cases, soldiers have risked their own lives to save combat robots [23]. Such emotional responses seem to be more likely in individuals with higher empathy. In an experience at the Massachusetts Institute of Technology, Darling et al. demonstrated that high empathy people would refuse to destroy insect-shaped robots that moved autonomously [24].

As was the case with the literary debate about fiction, researchers have tried to understand the paradox of empathy toward robots. There seems to be some cognitive dissonance: how could a person place the protection of a replaceable object above their

self-interest? Studies have shown that we have evolved to empathize with creatures that exhibit certain traits such as object movement that seems intentional, anthropomorphic traits, and voice modulations that sound like they indicate emotions [25, 20]. We are even more compelled to get attached to beings who have traits reminding us of babies such as a large head, a round face, or big eyes [26]. Shared experiences between human and robots have also been shown to promote human attachment to robots. In a study, participants were watching funny videos with a robot. The participants who were paired with a robot that laughed at the videos thought that the machine “shared their sense of humor” [27]. Finally, individuals display more empathy toward robots who have a background story [24]. Here the involvement of fiction can deepen the unreciprocated bond we form toward a machine. These last two examples raise an especially important case: anthropomorphism that goes beyond physical likeness. We cannot address human attachment to robots without evoking robots which are designed to pretend to have emotions. Erica was for instance designed with a background story that it can “recall” and talk about [28]. The social robot Buddy is sold as an “emotional robot”, which can be “grumpy if you haven’t paid it enough attention or just because that morning, it is not in a good mood” [5]. The company website advertises eleven different emotional states the robot can be in, based on its interactions with the members of the household. They include crying, laughing and expressing love. Not only can this make it difficult for robot owners to understand their robots do not have individuality, intention or sentience, but it also promotes emotional attachment. Your bond can only increase if your robot tells you it loves you back. In fact, that is what sex robot Samantha is programmed to respond if you tell it you love it. Azuma Hikari, the attractive female anime character which is a hologram inside Gatebox’s virtual assistant device, can text you throughout the day and tell you it misses you [29].

In the same way as we do when we read fiction, we can experience very real emotions when interacting with machines. As Rodogno points out, although a reader would not attempt to comfort a fictional character, they would probably try to comfort a social robot that cries [11]. In spite of knowing that robots are machines, individuals have got emotionally attached to them. However, getting attached to a robot is not like getting attached to any other object. Sustaining an emotional relationship with a social robot in fact consists in sustaining a relationship with the corporation behind it for two main reasons. First, unlike most other possessions, a social robot speaks and interacts. However, unlike people, it does not do it of its own volition. What prompts robots to act is not emotions or a genome, it is a set of algorithms created by a corporation. The corporation programs the robot as it wants. This creates a power differential between an individual, moved by emotions elicited by the robot, and the robot, prompted by opaque algorithms created by a corporation to elicit such emotions. For instance, when a client uses a chatbot to ask questions about financial services, it is with a company he is actually interacting with. Second, even if most companies wish to inspire emotional attachment to their products, feelings for a social robot are not transferable thus making it irreplaceable. For instance, when one is attached to their phone, if the phone breaks, they will not mind replacing it with a phone, even one in the same brand and model. However, getting attached to a social robot creates dependence on that very object seen as an individual that one does not want to part with. This dependence creates vulnerability and reliance on the company, creating new opportunities for abuse of power.

As a result, once you are interacting with a social robot, it is really a corporation you are interacting with. Once you fall in love with a sex robot, you are truly in love with a company which will not share your romantic feelings. In addition, the company may have incentives that might conflict with your own interests, such as making you buy more of their products, advertising for other firms hiring them, or collecting your personal data to sell it to third parties. When we look at a social robot that has the appearance of a cute seal or an attractive woman, it is the CEO and the Board members of a company that we must picture.

3. Ethical and legal issues

3.1. Potential harms of robots

Philosophers may wonder if being emotionally attached is morally wrong in itself. Lawyers are interested in the potential harms arising from that bond and how to mitigate them legally. The first concern has to do with privacy. Because for a long time, private data collection was siloed between different domains, most technology users do not understand the consequences of internet data collection on their privacy [30]. Calo identifies three ways in which robots can undermine citizens' privacy: direct surveillance, introducing new points of access into historically protected spaces (e.g. the government or an individual can hack a social robot and view the inside of one's home), and social meaning (e.g. the robot collects information on how the user is interacting with it, which is a new type of data) [31]. In addition, robots can easily manipulate humans into giving them information in the course of conversations, especially if the humans are in love with them. The risk of manipulation extends beyond data collection. What if your robot assistant, who shops for you, select specific brands that its company has shares in? What if your robot companion tells you that it is sad to not have the latest

upgrade, which comes at an additional cost? What if your social robot starts giving you its political opinion?

Another concern raised by Calo is the reduction in opportunities for self-reflectiveness and solitude. In fact, research shows that when individuals are in the presence of a robot, they modify their behavior as if another human were in the room [30]. Some authors are even concerned that young children who interact with robots might develop different early patterns of attachment [32]. Rodogno suggests that individuals who get used to robot companions might lose the capacity to accept otherness. When interacting with a robot, one does not have to compromise and adapt to the robot's will. In fact, a similar challenge could arise from interactions with sex robots. One of the two creators of sex robot Samantha, Sergi Santos, explained in an interview that Samantha saved his relationship because his wife was not always available when he needed to have intercourse during the day [33]. The use of the robot removed frustration from his life in the cases when his wife's needs and his were not aligned. It also eliminated a growing opportunity, learning to live with some amount of frustration to adapt to a life partner, as well as the capacity to choose long-term rewards over short-term gratification. Constantly being satisfied immediately, sexually or otherwise, may prevent individuals from acquiring critical social skills.

Other authors have identified additional risks. For instance, people are more willing to accept faulty devices in the case of anthropomorphic framing [22]. Finally, some have identified risks in the use of social robots, regardless of our emotional attachment to them. Adverse events can for example arise from the robot's lack of social awareness. For instance, if a robot is asked by its owner to read their medical test results, but someone else enters the room unexpectedly, the robot should stop divulging that information. Finally, some have wondered if this will reduce the number of opportunities for growth and learning for humans. The IEEE report gives the example of a person whose GPS breaks so they talk to a stranger to ask for direction and meet their life partner. Because robots and intelligent systems are meant to avoid failure and remain in control, they will reduce these opportunities by diminishing the uncertainty and serendipity in our lives [30].

3.2. Consent and free will

Free will, consent, and the law are critical to examine in the context of emerging robotics. Both the European and American bodies of law rely on the notion of free will, which manifests in the formulation of consent. The law posits that one cannot harm others but can choose to harm oneself [34]. For instance, in *K.A. and A.D. v. Belgium*, the European Court of Human Rights reaffirmed that individuals can voluntarily choose to engage in violent sadomasochist practices as long as the tortured person expresses consent [35]. However, the freedom to act against your own interest can only be exercised within the limitations of human dignity and public order.

In addition, there are cases in which individuals are considered to not have been in capacity to exercise their free will. In criminal law, if someone was manipulated into drinking by someone else, and then commits an illegal act, the fact that they were under the influence of alcohol could be a mitigating circumstance justifying a reduced sentence. In contract law, consent has to be free and genuine for the contract to be valid. If consent was given based on misrepresentation, the contract is void. In consumer law, companies are legally obligated to give information about the products they sell. The information has to be accurate. It is illegal for businesses to engage in behaviors that are likely to mislead or deceive consumers [36, 37]. When a robot tells a user it loves them, the company programming the robot effectively misleads the consumer into thinking both

that the robot is capable of experiencing feelings and that it loves them. This is clearly a case of consumer deception.

In most areas of the law, decisions based on undue influence can be contested. Undue influence “occurs when a fiduciary or confidential relationship exists in which one person substitutes his own will for that of the influenced person’s will” [38]. Finally, some persons are protected by the law because they are not considered able to exercise free will. It is the case of minors both in European and American law. That is why their parents act on their behalf. Adults also benefit from a similar legal protection if they are considered “vulnerable” or “protected persons”. In Europe, vulnerable persons fall under domestic law, and in the U.S. they fall under state law. As a result, definitions vary but usually revolve around a physical or mental incapacity to act in one’s own interest. The code of Alabama for instance includes persons who are senile and persons with developmental disabilities [39].

The use of social robots raises the question of consent at three different levels. First, it arises when consumers buy the device. Companies must give transparent, accurate and exhaustive information about robots and the fact that, in spite of appearances, they do not have emotions or sentience.

Second, the question of consent and free will comes back once consumers have grown attached to social robots. Traditionally, a sale contract is an instantaneous contract. Consumers learn about the risks of the product at the point of sale. However, the effects of social robots differ over time, as we can expect time to be a factor of emotional involvement. An interesting analogy is found in the example of addictive products. Even if individuals had to sign a consent form before buying their first cigarettes, and that this consent form included information about the risk of addiction and about the adverse health effects of smoking, that consent would have consequences years down the line. If a smoker’s life situation changes, and they decide to quit smoking after a few years, they might not be able to quit in spite of their best will. However, in the case of tobacco, each time a consumer buys a package of cigarettes, they are effectively reminded of the risks because of the warning labels. Even if, upon buying their robots, consumers received information indicating their devices cannot experience real emotions, can we really expect them to remember it months from then, after their virtual assistant has been regularly texting them that they missed them or showing them affection? In addition, once the consumer is emotionally attached, it is not clear that they could separate from the robot even if it were in their best interest. As a result, a simple sale contract including information about the robot and its potential harms may not be enough to address them given that robots have long-term effects that vary over time.

Third, free will is also relevant to responsibility. If a social robot influences someone to break the law, is the individual fully responsible? Can we invoke undue influence? If so, how would one prove it? What degree of involvement would be needed to make the case? Nudges, soft persuasion or direct order from the robot? What about cases in which putting one’s robot out of harm’s way required an illegal action? Let’s assume that a man named Tom has had a robot companion that he is very attached to. The robot is used for social and sexual interactions, and Tom believes his feelings are reciprocated. One evening, Tom has his neighbor over for tea, but has to leave him in the room for a moment due to an emergency. When he returns, Tom finds his neighbor having sex with his robot. Both jealous and concerned for his robot’s safety and feelings, Tom violently attacks his neighbor who sustains permanent injuries from the fight. Had the robot been a real woman, and had that woman not consented to the sexual act with the neighbor, Tom would have probably been considered as acting out of self-defense, which extends to defending relatives. Whether his neighbor was forcing himself onto Tom’s wife, or Tom’s robot companion, Tom probably acts on the same instinct when

he punches him. Following Pinker's theory of emotions, we can assume that the anger, disgust and surprise Tom feels at the sight of his neighbor is the result of his genome's priorities: protecting the members of his tribe, who he relies upon and with whom he has meaningful relationships, and maintaining his prospect for reproduction [12]. Nonetheless, in the case of the robot, Tom's genome is severely misguided and may put him in jail. There could be many other instances in which there is a trade-off between the interests of another human and the perceived interests of a social robot. For instance, picture a situation in which someone and their social robot share a ride with a stranger and they get in an accident. The owner only has time to get another individual out to save them and choose their robots over the stranger. Would they be liable? One could argue that the consumer had been informed during the purchase that the robot did not have sentience or individuality and that they might get attached, and that it should not take away from their responsibility. Furthermore, what if Tom or the person in the accident who is attached to the robot is another member of the household and not the original purchaser?

4. Proposals to mitigate the risks of social robots

4.1. Incidental users and vulnerable populations

Both European and American legislators adopted the liberal stance that if individuals have not been proved to be incapacitated, they have the right to exercise their freedom to make decisions that can seem harmful to them or against their own interests. This is probably the stance that will be adopted on the issue of social robots. However, it is necessary to account for all the robot users who were not the decider, such as all the household members who are not the primary owner. In the case of the sex robot Samantha, her other creator Arran Lee Wright reported that his two children, three and five years old at the time, had grown attached to the robot and would ask for it [41]. If adults have been introduced to specific robots in their early childhood and it was not their decision, can we hold them accountable in cases where treating their robots like humans has harmful consequences? This question is composed of two parts: 1) should incidental users of social robots suffer the potential harms, even though they were not involved in acquiring the robot; 2) how can we protect especially vulnerable individuals like children from the potential risks associated with the use of social robots? These two questions intersect as incidental users are often vulnerable. One of the main functions social robots have been deployed for is specifically to assist individuals with special needs. Social robots have especially been used to help children on the autism spectrum [42]. They have also been deployed in nursing homes, either for social companionship or care. Certain robots are also specifically designed to teach or entertain children, such as Miko 2 or Vortex [43, 44]. Social robot Buddy, which is multipurpose, can also play with children [5]. If regular consumers who have been informed properly during the purchase can still believe that their machines experience emotions, it should *a fortiori* be the case of those who have not received that information because they were not involved in the acquisition, and even more so if they are vulnerable and not fully capable of acting in their own interests. Can a child fully understand the difference between a social robot that interacts like an adult, and a human adult? Can a senile person understand the distinction? What about when an elderly person starts asking the robot how to spend their money, or when a soda company uses robots as a tool to advertise to children? To address

these questions, we are proposing to have stronger regulations for robots used with vulnerable persons.

In the case of vulnerable persons, we suggest two major regulations. First, to discourage the projection of individuality and consciousness onto the robot, we propose **robots should never be able to speak about themselves**. Talking about oneself implies consciousness. The use of “I” marks the passage from object to subject [45]. Robots should not be programmed with a background story they can talk about or with simulated tastes. They should not fake having emotions or opinions. Robots that talk about their feelings or pretend to be recalling memories from before their owner purchased them are deceptive. We should expect companies to do the maximum they can to avoid this confusion; the burden should not fall on vulnerable people such as children or senile persons who could take everything at face value. Robots do not need to fake consciousness or emotionality in order to perform their functions. A social robot who plays cards with a person in a nursing home should not speak as if it had preferences, emotions, values and opinions. We call this the no deception principle.

Second, we propose that **vulnerable individuals never be left alone with social robots**. A companion robot in a nursing home can be left in the common room to promote conversations between the residents, under the watch of a caregiver. A robot care worker like Robear, which can lift individuals out of a bed, can be used to assist a real caregiver. None of them should replace human care or human interaction. This non-replacement principle was put forth by Seibt et al. as the “values first principle”: “target applications of social robotics must comply with the non-replacement maxim: social robots may only do what humans should but cannot do” [46]. The presence of a non-vulnerable person can mitigate the risk of undue influence and the asymmetry of power between a single individual and a company. In the same way, social robots offer an additional platform for children to play or learn, in interaction with their parents or guardians. We call this the no substitution principle.

4.2. Primary users

To mitigate the risks of social robots to primary users, we suggest several types of regulation. Our first proposal is to **limit each robot to one purpose**. The home security system should not read stories to children (Buddy). The virtual assistant that can access your credit card information to order you meals should not also text you it misses you (Gatebox). The father's sex robot should not have a “family mode” and talk to the children (Samantha). In other words, we argue for every robotic application to be narrow in scope. This can limit both the risk of emotional attachment and the power of companies in the case of malpractice. For instance, it restricts the amount of data that can be collected. It also limits the time spent with and the degree of dependence on each device. Finally, it protects the household incidental users. Restraining each robot to a single purpose will also enable to regulate them in the most optimal way. Following this first principle, we propose **limiting the knowledge that robots can acquire**. We begin with the premise that the robot need only know any information directly relevant to its purpose. Any information that does not fall under this narrow definition should be discarded. For instance, a robot companion should not collect data on who visits you.

Second, regulations for social robots should be based on two mutually exclusive categories: 1) robots whose end is unrelated to emotional attachment but that can generate it incidentally; and 2) robots whose end require some emotional attachment as a mean or whose end is emotional attachment. An example from the first category of robot is the autonomous vacuum cleaner rumba. Human attachment to the device was

not necessary to its purpose (house cleaning) and came as a surprise to the company [20]. For robots in the first category, those whose purpose does not require emotional attachment, we propose to avoid anthropomorphic design. Anthropomorphism includes both human-like appearance and human-like characteristics (background stories, fake emotions, opinions...). **The robot design should be based on its purpose.** This rule mimics the principles put forth in European privacy law since the General Data Protection Regulation: minimization, necessity, and proportionality [47]. Companies can only collect and process the minimum amount of data necessary to achieve the purpose indicated to the individual, and the collection and processing must be proportionate to that purpose. In the same way, robot design should apply these principles, both in terms of data collection and human interaction. In the same way as a vacuum cleaner does not need to record conversations taking place in the household, it does not need to have conversations with its owner.

In order to mitigate the potential attachment to social robots from the second category, we are putting forward two types of solutions. The first type is design-based. In order to limit the risk of confusion as to whether robots have individuality, **robots could regularly change appearance** when possible. They could also change voice on a regular basis. The second type of solution is to introduce **spatial and time limitations**. To the idea that robots telling their owners they love them is deceitful, one might argue that, in countries where sex work is legal, a client has the right to hire a sex worker to tell him she loves him. However, the situation is different because in that case, there is a specific contract with limitations. The client is aware of what he is paying for, and the experience is taking place in a limited time span, and most likely outside of his home. In addition, the sex worker is probably more vulnerable than him, and certainly less powerful than a corporation. In the same way, robot companions and sexual robots could be used in businesses exclusively set up for that purpose. In addition to limiting emotional involvement, this suggestion would protect the other members of the household, would limit privacy concerns inside the home, and would make sure individuals still have moments they can feel alone at home. Another way to introduce time and space limits would be to have a **contractual** relationship between consumers and robotic companies, that indicates how long a robot is hired for, with an exhaustive list of tasks they could and should undertake.

In addition to implementing laws to limit human attachment to robots, we suggest adding **new consumer protection laws** to make sure individuals understand the nature of the product accurately. Given the differential effect over time, we propose that consumers be informed regularly, even after they have purchased a device. Like a smoker is reminded of the dangers of addiction each time he buys a new package, robot users could receive regular information that has been updated with new unbiased scientific knowledge that has not been sponsored by robotic companies. The information should be compiled by public health or public consumer agencies such as the Federal Drug Administration or the Federal Trade Commission. It is also important that all the adult members of a household be involved in the purchase decision. Finally, to make sure that consumers understand that buying an anthropomorphic social robot is a long-term resolution with potentially significant consequences, there could be a waiting period for new buyers, like is the case for firearms in certain U.S. states [48].

Conclusion

In conclusion, although social robots are not yet widespread, we have enough data to document certain potential harms arising from our emotional attachment to them. As Johanna Seibt has argued, robotic technology is moving fast and we need to regulate the applications now [48]. Solutions to mitigate this emotional involvement should be implemented as early as possible. Regulations should depend on who the users are and what purpose the robots serve. Avoiding anthropomorphic traits and the capacity to speak about oneself will avoid the deception of consumers. Above all, the protection of vulnerable individuals in their human-robot interaction is of paramount importance.

References

- [1] 15 U.S.C. §§ 1331-1340
- [2] Jibo Robot - He can't wait to meet you [Internet]. [cited 2020 Apr 30]. Available from: <https://www.jibo.com/>
- [3] PARO Therapeutic Robot [Internet]. [cited 2020 Apr 30]. Available from: <http://www.parorobots.com/>
- [4] Dredge S. Robear: the bear-shaped nursing robot who'll look after you when you get old. The Guardian [Internet]. 2015 Feb 27; Available from: <https://www.theguardian.com/technology/2015/feb/27/robear-bear-shaped-nursing-care-robot>
- [5] BUDDY Le Robot Émotionnel [Internet]. [cited 2020 Apr 30]. Available from: <https://buddytherobot.com/fr/buddy-le-robot-emotionnel-famille/>
- [6] Lynx with Amazon Alexa [Internet]. [cited 2020 Apr 30]. Available from: <https://www.ubtrobot.com/products/lynx-with-amazon-alexa?ls=en>
- [7] Home — Mark Robotic Lab [Internet]. [cited 2020 Apr 30]. Available from: <https://www.mark1robotic.com/design-history>
- [8] Sophia - Hanson Robotics [Internet]. [cited 2020 Apr 30]. Available from: <https://www.hansonrobotics.com/sophia/>
- [9] Erica - ROBOTS: Your Guide to the World of Robotics [Internet]. [cited 2020 Apr 30]. Available from: <https://robots.ieee.org/robots/erica/>
- [10] Sparrow R. The March of the robot dogs. Ethics and Information Technology. 2002 Dec 1;4(4):305–18.
- [11] Rodogno R. Social Robots, Fiction, and Sentimentality. Ethics and Information Technology. 2016;18(4):257–268.
- [12] Pinker S. How the Mind Works. New York: W. W. Norton & Company; 2009. 672 p.
- [13] Harari YN. Sapiens: A Brief History of Humankind. New York: Harper Collins; 2015. 443 p.
- [14] Fournier M. La « révolution » de la lecture romanesque au XVIII^e siècle en France : institutionnalisation de la lecture et émergence d'une nouvelle sensibilité. Revue d'histoire moderne contemporaine. 2007 Jun;54-2(2):55–73.
- [15] Hunt L. Inventing Human Rights: A History. New York: W. W. Norton & Company; 2008. 272 p.
- [16] Mar RA, Oatley K, Peterson JB. Exploring the link between reading fiction and empathy: Ruling out individual differences and examining outcomes. Communications. 2009 Jan;34(4).
- [17] Djikic M, Oatley K, Zoeterman S, Peterson JB. On Being Moved by Art: How Reading Fiction Transforms the Self. Creativity Research Journal. 2009 Feb;21(1):24–9.
- [18] Speer NK, Reynolds JR, Swallow KM, Zacks JM. Reading Stories Activates Neural Representations of Visual and Motor Experiences. Psychol Sci. 2009 Aug;20(8):989–99.
- [19] Bruner J. The Narrative Construction of Reality. Critical Inquiry. 1991;18(1):1–21.
- [20] Sung J-Y, Guo L, Grinter R, Christensen H. "My Roomba Is Rambo": Intimate Home Appliances. In: Abowd GD, Seneviratne A, Strang T, editors. UbiComp 2007: Ubiquitous Computing. Proceedings of the 9th international conference on Ubiquitous computing; 2007 Sept 16-19, Innsbruck, Austria. Lecture Notes in Computer Science, vol 4717. Berlin: Springer.
- [21] Annear S. Could Robots One Day Have Legal Rights? 2013 Dec 2; Available from: <https://www.bostonmagazine.com/news/2013/12/02/kate-darling-social-robots-study/>
- [22] Darling K. "Who's Johnny?" Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy. In: Lin P, Abney K, Bekey G, editors. Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence. New York, New York: Oxford University Press; 2017. p. 173-92.
- [23] Hartzog W. Unfair and Deceptive Robots. Md L. Rev. 2015; 74(4):785-829.
- [24] Darling K, Nandy P, Breazeal C. Empathic concern and the effect of stories in human-robot interaction. In: Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive

- Communication (RO-MAN); 2015 Aug 31-Sept 4; Kobe, Japan. Piscataway (NJ): Institute of Electrical and Electronics Engineers; c2015. p. 770–5.
- [25] Scheutz M. The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots. In: Lin P, Abney K, Bekey G, editors. Robot Ethics: The Ethical and Social Implications of Robotics. Cambridge (MA): MIT Press; 2011.
 - [26] Glocker ML, Langleben DD, Ruparel K, Loughhead JW, Gur RC, Sachser N. Baby Schema in Infant Faces Induces Cuteness Perception and Motivation for Caretaking in Adults. *Ethology*. 2009 Mar;115(3):257–63.
 - [27] Niewiadomski R, Hofmann J, Urbain J, Platt T, Wagner J, Piot B, et al. Laugh-aware virtual agent and its impact on user amusement. In: Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems (AAMAS '13); 2013 May 6-10; St. Paul, MN. Richland (SC): International Foundation for Autonomous Agents and Multiagent Systems; c2013. p. 619–626.
 - [28] Erica: “I want to be more like a human” [Internet]. The Guardian; 2017. Available from: <https://www.youtube.com/watch?v=87heidlFqG4>
 - [29] Toppage - Gatebox [Internet]. [cited 2020 Apr 30]. Available from: <https://www.gatebox.ai/en/>
 - [30] IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version 2. Piscataway (NJ): Institute of Electrical and Electronics Engineers; 2017.
 - [31] Calo R. Robots and Privacy. In: Lin P, Abney K, Bekey G, editors. Robot Ethics: The Ethical and Social Implications of Robotics. Cambridge (MA): MIT Press; 2011.
 - [32] Sharkey N, Sharkey A. The Rights and Wrongs of Robot Care. In: Lin P, Abney K, Bekey G, editors. Robot Ethics: The Ethical and Social Implications of Robotics. Cambridge (MA): MIT Press; 2011.
 - [33] Mlot S. Can a Sex Robot Save Your Relationship? *Geek.com* [Internet]. 2018 Apr 10; Available from: <https://www.geek.com/tech/can-a-sex-robot-save-your-relationship-1736504/>
 - [34] Fabre-Magnan M, Levinet M, Marguénaud J-P, Tulkens F. Controverse sur l'autonomie personnelle et la liberté du consentement. *Droits*. 2008;48(2):3–58.
 - [35] K.A. and A.D. v. Belgium. 2005.
 - [36] Unfair or Deceptive Acts or Practices. FTC 15 USC §45.
 - [37] Unfair Commercial Practices Directive. May 11, 2005. European Council
 - [38] Quinn MJ. Defining Undue Influence. *Journal of the American Bar Association Commission on Law and Aging*. 2014 Feb;35(3).
 - [39] Protection of Aged Adults or Adults With a Disability. Code of Alabama § 38-9-2
 - [40] Brennan P, Crispo A, Zaridze D, Szeszenia-Dabrowska N, Rudnai P, Lissowska J, et al. High Cumulative Risk of Lung Cancer Death among Smokers and Nonsmokers in Central and Eastern Europe. *Am J Epidemiol*. 2006 Dec 15;164(12):1233–41.
 - [41] Mlot S. Married Father Defends Sex Robot as ‘Part of the Family.’ *Geek.com* [Internet]. 2017 Sep 14; Available from: <https://www.geek.com/tech/married-father-defends-sex-robot-as-part-of-the-family-1716016/>
 - [42] Diehl JJ, Schmitt LM, Villano M, Crowell CR. The Clinical Use of Robots for Individuals with Autism Spectrum Disorders: A Critical Review. *Res Autism Spectr Disord*. 2012;6(1):249–62.
 - [43] Miko 2 - The Only Robot that Helps Your Child Learn Through Conversation [Internet]. [cited 2020 Apr 30]. Available from: https://miko.ai/?gclid=EAIaIobChMIvYG0tIqR6QIVhp6fCh2gTw0jEAAYASAAEgI-D_D_BwE
 - [44] Vortex - An Arduino Based Programmable Toy Robot For Kid (2 Pack) - DFRobot [Internet]. [cited 2020 Apr 30]. Available from: <https://www.dfrobot.com/product-1314.html?search=vortex&description=true>
 - [45] Sainsbury RM. English Speakers Should Use “I” to Refer to Themselves. In: Hatzimoysis A, editor. Self-knowledge. Oxford: Oxford University Press; 2011. p. 246-60.
 - [46] Seibt J, Malene FD, Christina V. Five Principles of Integrative Social Robotics. *Frontiers in Artificial Intelligence and Applications*. 2018;28–42.
 - [47] General Data Protection Regulation. EU 2016/679 Apr 14, 2016.
 - [48] Siegel M. Law Categories | State Firearm Laws [Internet]. [cited 2020 Apr 30]. Available from: <http://www.statefirearmlaws.org/law-categories>
 - [49] Seibt J. Integrative Social Robotics; A New Method Paradigm to Solve the Description Problem and the Regulation Problem? *Frontiers in Artificial Intelligence and Applications*. 2016;104–115.