University of Twente

Exploring Human-Robot Social Relations

by

Stefan Weijers

s0203769

A thesis submitted in partial fulfillment for the degree of Master of Science

in the

Faculty of Behavioral Sciences

Philosophy of Technology

Department of Philosophy

First Supervisor: Johnny Søraker

Second Reader: Mark Coeckelbergh

October 2013

University of Twente

Abstract

Faculty of Behavioral Sciences
Philosophy of Technology
Department of Philosophy

Philosophy of Science, Technology and Society

by Stefan Weijers
s0203769

An age where robots and advanced intelligent agents are all around us is nearing. It is important we explore the ways we will relate to these future robots, because that way we can try to avoid certain scenarios while encouraging others. The 'lovotics' research domain sketches a future where robots can be our friends, lovers and family, but is that really a possibility? To find the answer to that question we need some way to identify types of social relations of humans and robots.

Therefore I start with defining a social-relation framework that can differentiate between human social relations such as friendship, family relations, business relations and many others. Before we can apply this framework to human-robot relations however, we need to explore the capabilities of future robots and identify to what extend these robots can have the required properties for social relations.

Finally we will see what shortcomings robots necessarily have that will limit our social relations with robots to a certain level.

Acknowledgements

Many thanks to Johnny Søraker, my supervisor, for giving me advice on the literature and feedback on the drafts. Also thanks to Laura van der Straaten, Corrie Weijers and Rob Weijers, for proofreading my thesis and providing the suggestion to use a kiviat diagram to represent the framework.

Contents

Chapter 1

Introduction

1.1 The robotic near future

Robots and artificial intelligence have been an area of interest since even before the first electronic computers were introduced. Many saw a threat in a potential rising of robots and see reason to restrain robots by laws for the safety of humans (Asimov and Reilly, 2008). Others see potential in robots as agents that give care and help cope with the increasing percentage of elderly people (Sparrow and Sparrow, 2006). Robots are supposed to become almost indistinguishable from humans, and friendship with robots would flourish.

In Hollywood-media there are numerous science fiction scenarios for a positive and negative future in regard to robots. In 'Robot and Frank' (2012) a care robot butler is used by Frank to get back into the habit of stealing. The robot seems to understand spoken language perfectly and can learn new physical activities like picking locks. However, his moral judgement about stealing seems lacking. In '2001:A Space Odyssey' (1968) an intelligent computer tries to kill the humans aboard the spaceship in order to fulfill its mission successfully. This artificial agent is also capable of understanding language (it can even read lips) and knows how to deceive and manipulate people.

The are numerous movies and stories with a plotline revolving around very intelligent artificial agents and robots of some sort. From the 'Hitchhikers guide to the galaxy' to 'Terminator' the role of robots of some sort in our near future seems undeniable to science fiction writers. There are however also less fictional examples that seem to indicate that robots will have a huge impact on the social lives of many people in the near future. Especially in the domain of caring for the elderly. For example the Mobiserv [1] is a

[1] http://www.mobiserv.info/

robot designed purely for social interaction and taking care of people suffering from mild dementia. Furthermore there is a whole research domain called 'lovotics' that researches robots in order to have a love-like relationship with them. (Samani et al., 2011) A future with robots might seem inconceivable to many, but the technological advancements in the area of robotics might considerably change society in the near future. It is no longer science fiction, there will be social robots in our homes in the near future. An important question is how our social relations will be impacted by these social robots. Will robots replace lovers, friends and family or will a special type of human-robot social relation emerge?

The 'near future', a term used many times in science and also in this thesis, but what does it really mean? The way 'near future' is used in this thesis is as a timespan from now till anywhere in this century and perhaps a bit beyond. This seems to be a quite long time, as it is more than most of us will live to see, but it is relatively near in terms of human history. That said, the pace at which computers, cars and planes have developed in the past 100 years is astounding. Therefore it seems that many new technologies will be possible after another such 100 year timespan. An important point to note is that the near future does not account for wars, epidemics or other global disasters of any kind. The point of this setting is that there should be significant time, resources and effort available for research and production so that the best effort of the global scientific community is capable of achieving one certain goal. In this thesis, the goal for that near future is the most intelligent social robot ever constructed. If all scientists could work on this robot for a significant amount of time, what is the best they could possibly create? This is the mindset that near future is all about, it is not about predicting the future, but rather seeing what possibilities and more importantly, impossibilities there are. Questions we should discuss include; what kind of social relations could we possibly have with such advanced robots or artificial agents? Will these artificial agents have any form of feelings or emotions? Or will they just be mindless agents that do not care about our well-being at all? Not just any type of robot is in the realm of possibility for the near future. We need a way to separate the science from the fiction.

1.2 Research question

One of the most important questions that can be raised is whether the social relations that are formed with these fictional robots from the near future are really friendships or whether they are social relations of a different kind. A better question would be whether such future robots could really exist. If not, what kind of robots will exist? What can our social relation be to the real near future robots?

Samani (Samani and Cheok, 2010) tried to reproduce 'love' in a mathematical way, so that his sponge-like robot could change his behavior to simulate 'love'. Although an interesting approach it does not seem that this method produces a definition of 'love' and 'friendship' that is close to the actual human experience of such emotions. Surely the robot sponge might induce a sense of feeling in the human being, but this is not necessarily the same kind of love or affection we would have for another human being. To properly take the complexity of human social relations in mind we should not rely on mathematical formulas for an approximation of human-robot relations, but we should rather treat the way humans and robots relate as a philosophical question and explore it that way.

The main research question therefore is: How can we describe human-robot social relations? To answer this question we need to answer the following sub questions:

- What are social relations?
- How can we compare social relations?
- What are the near future capabilities for robots and artificial agents?
- How do we determine these capabilities?
- What role does consciousness play in social relations?
- To what extent can appearances play the same role as reality?
- How can we map robot capabilities to properties of social relations?

The method for this research is based on a literature research that combines three domains. The first domain is philosophical in nature and is about friendship. This philosophical inquiry into friendship touches upon many philosophers from the past to get an understanding of what friendship consists of and how friendship relations are formed. The choice to base social relations upon friendship theory is based on the premise that friendship is one (if not the) of the most important social relations. This will be the subject of chapter 2. The end result of this inquiry is a social relation framework that will later be used to identify the types of social relations applicable to human-robot relations.

The second domain is computer-science in nature and is about the capabilities of near future robots. This domain combines some literature from social studies of technology to find a method to make proper judgement about the capabilities of technologies in the future. Any speculation about the future is uncertain to some degree. Using proper

methods helps to reduce the uncertainty and allow for better and more useful predictions. The other half of the research in this chapter is about the necessities and capabilities of robots, artificial agents and where the difficult problems of constructing proper intelligent agents lie. These points will be discussed in chapter 3 and the goal of that chapter is to show that the current and near future capabilities are not so fantastic after all, because the problems for capabilities like emotion, thought and understanding are very difficult to program and control.

The third domain is again a domain in philosophy. The domain is about the way we understand the world and ourself. In this literature research we explore the differences between appearances and reality and how that matters. This is of great importance when we speak of consciousness, as many arguments against conscious robots are of a biological nature. If robots could be conscious this would be a big step in solving the main question, as a social relation between humans and other conscious beings (artificial or not) seems less controversial. However we will find out that the arguments against robot consciousness are not biological in nature but rather about complexity. Chapter 4 will expand on this part of the research and its goal is to show how one can think of robots and the way they can appear to us from the perspective of a computer scientist.

Finally these three domains are combined. The capabilities and the way they appear are tested against the framework for social relations to show that friendship with robots is not the way we can describe such a relationship.

This is the way sub questions will be discussed in the chapters to come.

Chapter 2	What are social relations?
	How can we compare social relations?
Chapter 3	What are the near future capabilities for robots and artificial agents?
	How do we determine these capabilities?
Chapter 4	To what extent can appearances play the same role as reality?
Chapter 5	How do the capabilities of near future robots map?

Chapter 2

A Framework for Social Relations

2.1 Introduction

One of the significant differences between humans and animals is the fact that we are extremely social animals. We have language, social systems and various technologies with the only purpose to communicate. Whether social relations have an intrinsic or extrinsic value for humans is under debate. According to Fiske (1992) a number of psychologists argue that humans are inherently asocial, and that we merely engage into social relations in an instrumental way. Social relations are there in order to achieve some other purpose, for example happiness or material advantages. While Fiske (1992) argues that humans are inherently social beings, and that the organisation of our lives fundamentally revolves around our social relations. Whichever point of view one wants to take, it seems beyond dispute that social relations are a valuable addition to our lives.

One of the most obvious and fundamental social relations is that with family. A family often consists of a multitude of different relations. These relations are mostly hierarchical in nature - Think for example about father-son or older-brother-younger-brother relations. Other common relations people have include colleagues, business relations, the relations customers have with the shopkeepers, one's significant other and of course friendship.

Friendship is perhaps the most interesting relation of all. Cicero would argue that without friendship life would not be worth living. One of the core questions in many of the philosophical works regarding friendship is "why would we engage with other people at all?". It is in the analysis of friendship relations that much of the other social relations humans engage in find their place.

The first half of this chapter explores the different aspects of social relations. Including the motivation, symmetry and relations to non-humans. In the second half, we will use these aspects to define a framework that can map social relations. The core thesis of that framework is that all social relations can be described as a configuration of four properties.

2.2 Aspects of social relations

2.2.1 Motivation

Let us begin with one of the core questions the philosophical tradition about friendship; "why would we engage with other people at all?". Aristotle argues for three different motivations for social relations. Firstly for the goodness in someone, secondly because they are useful to us, and thirdly because of pleasure.

According to Aristotle (Aristóteles, 1991), only friendships based on the goodness in someone's personality can be considered "true friendship". Aristotle writes that a real friendship is reciprocated goodwill (NE vii, 1155b), which means that only two persons that consider each other as a good person can be true friends. In contemporary philosophy, this concept of goodness is hard to apply to modern social relations. We could try to understand it as a social relation "for no good reason" or "just because". This would mean that the question; "why are you friends?", in the case of true friends, should result in a nonsensical answer of the former sort. Aristotle has a different take on this goodwill. In Aristotle one should relate this goodwill back to himself. One should care for his friend, for the friends sake, as a friend is a manifestation of one's own values and virtues. Through a friend one can shape oneself since the very virtues that one finds important in friends are the virtues one thinks oneself ought to have as well.

A second motivation, according to Aristotle, is that of usefulness. We can relate to another because he or she is useful to us. A good example of this kind of social relation might be a business relationship. A relation where both parties have a mutual interest in each other for their own sake. In this example the usefulness motivation seems quite clear. However things get more complicated when we take a closer look at usefulness. We could also imagine a relationship about usefulness that is non-symmetrical. For example non-symmetry would occur when one has significantly more resources than the other, and that the other is dependant on the first for his or her food and income. Even in the modern world many relations are based on this usefulness motivation, for example in the professional world, between bosses and employees, to fellow colleagues and customers.

Even in non-professional settings, such as between parent and child, although it is harder to speak of a certain 'motivation' in the case of family.

The third motivation Aristotle mentions is that of pleasure. A social relation purely based on pleasure is an interesting idea. An extreme example of this would be a sex addict, that just seduces one person after the other only for physical pleasure. A more common example would be a "buddy" (authors terminology), who is not actually a close friend, but the relationship is similar to one when engaging in a particular shared activity. Take ones hobby, sport or game and the relation with fellow practitioners is that of pleasure. Naturally some of them can become more close friends, but with many all contact would be lost if one were to quit the shared-activity.

This separation of motives is a useful one, because it helps segregate some of the different relations. The subjacent paragraphs will use this analysis in a slightly altered way as the sharp distinction between the cases seems too simplistic. Modern friendship is not motivated in this way, but is rather a combination of these motives and some external activity. Like C.S. Lewis's (Lewis, 1960) observation that one of the differences between lovers and friends is that lovers talk explicitly about their relationship (i.e. tell the other that you love him or her) while friends do not talk about their friendship at all. Rather friends talk about the subject of their relation, a shared interest or activity. This also means that friends do not need to explicitly talk about or understand each others motivations for engaging in the friendship. This observation also works for all other relations, like business or family relations. One doesn't talk about how great their business relation is, but rather about business. Similarly, one does not talk about family as family, but rather about ones life in the family.

Motives as described by Aristotle only get questioned when the relationship is not working out. The question why one's business associate is still a business relation is only asked when it is no longer mutually beneficial. Similarly the question why one's buddy (pleasure relation) is still one's buddy only comes up when at least one of the persons no longer finds pleasure in the shared activity. To continue relations with these problems a different purpose or interest must be found. Business relations or buddies can turn into friendships by adding new shared activities or interest to the relationship - Therefore widening the domain of the relationship beyond its original motivation. That is how we create friendship relations according to Lewis. Pleasure or usefulness is no longer the only motivation of a social relation. This is also why questioning motivations of a social relation does not really make sense, as when the question arises, the motivation for the respective social relation is already lost.

The love relation is different than the others in this regard. Unlike friendship and the other types of social relations lovers indeed do talk explicitly about their relationship

to each other. This would mean that the "love" social relation is indeed different and distinct from the other social relations, but not mutually exclusive. One's greatest achievement is being friends with one's lover, because not only is there physical attraction towards each other, there is also a rich mutual base of shared interests and activities that strengthen the relationship beyond physical attraction.

2.2.2 Asymmetrical relations

So far the social relations being discussed were mostly symmetrical in nature. A friendship relation is symmetrical for example, as it is the mutual motivation and shared activities that count. Philosophers are quite explicit about this symmetry requirement. Socrates, for one, asks whether it is necessary for both persons to love (In modern terms "have affection" or "care for") each other in order for there to be friendship (Pakaluk, 1991, Lysis:212-213). In an extreme one sided social relation where person 'A' loves 'B' while person 'B' hates 'A', it makes no sense to wonder who is the loved one and whom the hated, as both persons would perceive the other as both a friend and a hated person (let us say, enemy). This is because one either does "not-receive" love or "not-give" love (a mark of an enemy) and one "receives-love" while the other "gives-love" (the mark of a friend). In a sense these persons have an extreme asymmetrical relation, but at the same time perceive the other in the same way. Namely, as an enemy-friend. A social relation like this seems a bit far fetched, but asymmetrical relations do exist.

Deceit is a clear cause of asymmetrical relationships, as one of the persons deliberately gives a wrong impression of the relationship. It is most likely that the relation ends as soon as his or her goal is reached. We could also imagine asymmetrical relationships without any ill-will. Asymmetrical social relations resembling friendship seem rather common. For example a "friendship" where one is way more interested in the shared activities than the other. Another option is a social relation in which for one of the persons the presence of the other can annoy him or her more than reversely. Perhaps the most obvious kind of asymmetrical social relations is that of one-sided romantic love. Interestingly in the case of one-sided romantic love it is possible that the receiving party of the love has no idea that this is the case, or in the most extreme case doesn't even know of the existence of the giver. In real friendship relations this is quite different, as it seems to be a relation that needs to grow. A social relation needs the attention of both persons. In that light, it is quite astonishing that one-sided "friendships" (the quotation marks because one would not call this real friendship) can exist, as it would require one of the two to constantly compromise or subdue his personality to tolerate the other. It seems quite reasonable to ask, why would anyone do that? Perhaps for secondary motives (utility), or perhaps he or she doesn't dislike the other totally, but

rather partially, at certain moments, while liking his or her company at other times. One might ask oneself, how many of your friends can you stand being with for whole days without them or the activity getting annoying at some point?

An asymmetrical relation of a different kind is that between man and animals. A dog is said to be mans best friend. An interesting claim, considering that ancient friendship theory requires equality and and self-sufficientness from both entities. Even if these aspects are clearly not present there is something to say for this relation as it seems that dogs do show unconditional affection, trust and engage in shared activities. All of which are necessary conditions for a friendship but not sufficient according to the definition of friendship by Aristotle. That said, it seems a clear example of an asymmetrical social relationship, but what makes it asymmetrical? Is it that humans are intellectually superior to dogs? Or because we 'own' dogs? I doubt the dog is aware that it is 'owned'. And intellectual superiority seems a dangerous path to go as vital reason for the relation to be asymmetrical, as it might entail that (sufficiently) smarter persons can never have a true friendship with (sufficiently) less intelligent persons. Perhaps only if the smarter keeps the discussion on a level the other can understand.

Interestingly enough, philosophers on friendship do not really touch the subject of intelligence and the effects a big difference in intelligence can make regarding friendship relations. Examples of true friendships (such as in Plato, Montaigne) always seem to occur between persons of not only rather equal, but also high intelligence and social standard. This could be because those are the ones in history that would be more prone to find the time to write about their friendship - Even though true friendship seems rather elitist at times. Terms like 'common' (Montaigne, 1958) friendships suggest not only that they are more abundant, but also held by the common people, the non elite. Montaigne argues that these common friendships are motivated by services and benefits, and therefore of a lesser kind.

What about relations between good and bad people? Can they exist? Plato asks this question as well. He assumes that it is impossible to have a true friendship with a bad person, because bad persons would do injustice to others (good and bad), whom could therefore never be true friends. So only the good and the good would be able to be friends, but this seems impossible as well. A good person would not prize any other person, because he is fully self-sufficient (a strange way to define a good person for contemporary philosophers). Thus he concludes that only the neither-bad, nor-good persons want to be friends with good persons, because they are attracted to them out of need. This need is created by the existence of bad persons. The illustration Plato uses is that the body (with is neither good nor bad) is in need of medicine (the good) because of the existence of sickness (the bad) (Plato, Lysis: 217). This conclusion is

false according to Plato, because there would still be desires for friendship even if the bad would not exist. Therefore it cannot be on the account of the bad that friendship with the good exists. Plato leaves it at that. It must be more than just likeness[1] and goodness, but there is also an innate desire for friendship in mankind.

This innate desire might be related to why we relate to others for "no good reason". Not only does this desire cause friendships, it also seems part of the asymmetrical relationships. Asymmetrical relationships often seem easily accepted, without reasoning whether or not tolerating the asymmetry is in the best interest of both.

2.2.3 Inanimate Objects

So far we mainly talked about social relations with other humans. We also have a kind of relation with objects however. What is your favorite object? Most likely some object you have emotional attachment to, a gift from a friend, a memento of a lost family member. Or perhaps it is an item you use from day to day such as your phone or a piece of clothing. Even living things like plants could be someone's favorite object. What differentiates certain objects from others? How could we describe our relation to objects?

We can start this inquiry by wondering why we prefer certain objects over others. Norman in his book "Emotional Design" (Norman, 2007) writes about how we get attached to certain objects. One of the first rules is that there is no single design that everyone likes. People have their own preferences and concepts of beauty, influenced by their culture, and have different needs of an object.

A way to see objects is that objects are purely instrumental to our will. This would mean that all objects have as their primary reason of existence (raison d'être) the fulfillment of their intended purpose. This has as logical consequence that either all objects do not favour any special attention from us apart of their use, or the objects that are our greatest favorites are the ones that perform their functions the best. Both of these instrumental views however completely disregards emotional value in objects.

Norman describes that also with objects we have a form of bonding, a newly bought knife might feel weird and unfamiliar at first use, but as time goes by and the object gets more familiar we feel at ease with them. Even though the knife might not be the best knife around, we still prefer to use it because it is familiar to us. A striking similarity to how we treat other people.

[1] Likeness is about similarity and aspects persons have in common. In this case being good or bad.

Function is an important aspect of an object. If an object fails to fulfill its requested functions properly it is hard to love it or bond with it. Notice that this even holds for objects whose primary function is that of invoking a memory. A memento of a lost family member works best if it actually helps you to remember that person, so it should have some property that distinguishes it from other objects. It should be something related to that family member, but also not be too common. Inheriting a single Ikea plate, that is exactly alike your other plates will make it harder for it to fulfill its function as a memento. Either none of the plates invoke a memory or all of them will, regardless of how well the plate itself functions as a plate. Indeed, it might be better for its function as memento for it to be a non-functional plate (for example damaged or engraved) as that gives it some personal touch.

According to Norman, a way to increase the bond with inanimate objects is customization. In a sense customization is giving an object the status that it is unique or specialized to fit one person in particular. This means that such object cannot be replaced easily and seems to become more important. In friendship we can find similarities to the concept of customization. In friendship we can often speak of "I know him like no one else does" or "You are the only one I can talk to". Friends in that sense are "customized persons" suited to fit our social relation.

Beauty is a concept which explaining could cover whole books. "What is beauty?" and "What does it do to us?" are both important questions. For our purposes it is sufficient to argue that the appearance of an object influences our attraction to it. We are more likely to buy things that are pretty, even if they are less functional than ugly things. Surely there are tasks where form follows function. Where it is most important that some object does its job well, no matter how it looks. Very often however, there are situations where function is not always the most important. Norman, in his book Emotional design, begins his story by telling us about his three special teapots. Two of which have interesting designs that catch the eye, but are not all that functional. One of these teapots has its handle on the same side as its spout, effectively causing the user to be extremely uncomfortable while using it. Another teapot has an interesting three step process for brewing tea that includes putting the pot on its back, its side and normally, reflecting the stages in the teas darkness. Tilting it brings the tea leaves, that reside on an internal shelf, in contact with the water. When upright, the leaves do no longer touch the water and thus prevent the tea from becoming bitter.

Both these designs catch the eye of visitors and give the owner a story to tell about these teapots. The point Norman tries to make with these teapots is that people can love their belongings even if they are not functional. What such object needs is that it

is special, fun to use or to look at, and stays that way for a long time. We do not only love objects for their functionality, but also for their beauty or intelligence in design.

Art is another example of beauty without 'function'. Art can be interesting, art can be engaging and art can even have educational, relaxing or emotional functions. One might wonder how something without a function can have these educational, relaxing or emotional functions. The point is that a piece of art does not have a physical function such as a hammer, which physically helps us putting nails in wood, or a refrigerator which physically keeps its contents cold. Naturally all these objects might also have non-physical functions. For example the hammer might remind you of your grandfather, who is no longer around as he used to own it. These non-physical functions are the memory inducing kind of functions. Art is especially prone to have some sort of emotional effect on some persons. It might be awe for the skill of the artist, or the sheer size of the object, but also the unsettling scenery depicted. In historical art we can also discover a part of our past, which makes these works interesting. Some might not even be considered art at that time, such as the personal portraits, but through the passing of time our relation with these objects has changed.

Another interesting aspect of liking objects is that it is so easy to share this love by showing it or telling a story about it. Unlike in eros, sexual love, the love that is focussed upon pleasure of a one single entity and does not warrant being shared at the risk of jealousy (Lewis, 1960). Sharing the pleasures induced by objects is quite natural. Just look at the many museums that display loads of objects for people to appreciate. It would be especially fortunate to find someone who can passionately talk about these objects and shares the pleasure he attains from it. There is no jealousy in sharing the pleasures of objects, there is no jealousy in the sharing the knowledge. Jealousy in eros is one of the most important emotions, one that does not seem to exist in admirations. We all accept that we have preferences for certain activities or objects, and instead of being jealous when finding someone with similar interests, we rather have these shared interest become a part of the bond between friends. Shared interests for 'objects' such as engines, planes, art, sewing, fashion, gaming, food, music and many more, can all be the basis of a strong friendship. The kind of love present in admirations of different entities, persons or each other without jealousy is an important aspect of social relations.

Lewis argues that relations for pleasure and business are what he calls companions. It is a valuable relation as well, but different from friendship. He writes: "Companionship is only the matrix of friendship." Friendship arises from companionship, as two companions can discover they have mutual interests and tastes others do not. Wanting friends does not work - As friendships always are about something. Friends have an intention, a common interest or taste. Without any of that, there would be nothing for the friendship

to be about. Thus if we take Lewis seriously, this would mean that if someone has no passion or love for some object, music or activity, he or she is unable to form a friendship with anyone.

2.2.4 Entity Differences

Human-dog relations were mentioned briefly before in the section about asymmetrical social relations - In this section we focus a bit more on the entity differences. In asymmetrical social relations the relation is not equal, thus one party has more power, rights or resources than the other. This could result from unequal entities (such as a dog-human relation), but this is not necessarily the case. When we look at a parent-child relation for example, we see that even though the relation is asymmetrical, both should still treat each other with the respect befitting of humans. This means that even though the relation is asymmetrical, the entities can be considered equal in worth or physical and mental capabilities. The asymmetry in these cases comes from a difference in resources, rights and power that comes with a certain social status.

Human relations to for example a dog is an interesting one. A dog generally has less capabilities than a human. Note that I do not want to compare the best dog to the dumbest human, as that dog would most likely be smarter and have more capabilities (the argument of marginal cases). This argument is not about drawing borders between the capabilities, but rather resolves from the premise that dogs are in general, less capable, than humans when we speak of communication, learning, and the complexity of the tasks they can perform.

In a sense the difference between humans and dogs would mean that genuine friendship is impossible, as a dog and a human cannot have a mutual intention or interest that binds them. Although if one compares typical dog behavior with human behavior it is clear that these behaviors are directed from and to the dog. Similarly to a love relation, the social relation between dog and owner seems directed towards each other only. Therefore it would seem that this pet-relation is closer to a love relation than to friendship.

Love for inanimate objects and the love for animals are quite different however. Where love for inanimate objects is for their beauty and/or function, love for animals seems more genuine, more like human to human. A condition for this however, is that the animals are able to "love back". Some persons keep spiders as pets, and they might love them, but I doubt spiders have the capability to love their owners, and I think this shows in the human-pet relation as well. If words like beauty, intriguing, thrilling or exciting seem to describe the relation one has to his or her animals, it seems closer to the love of an inanimate object. While if words like trust, interaction and caring seem to describe

it better, it is most likely a pet that seems to reciprocate the love and attention, a pet that seems to be aware of its relation to you.

Now if we take these insights back to simple human-human social relations we can see that preferably true friends match all of these words. Trust, interaction and caring do not seem to warrant much explanation. Intriguing, thrilling, exciting and perhaps most of all beauty might. Intriguing, thrilling and exciting seem states of mind that are quite temporary, so why are they part of a lasting social relation like friendship? Friendships thrive on doing exciting or thrilling things together. Perhaps they become more thrilling or exciting because they are being done together. Without stories, shared activities or otherwise exciting ways to spend time with friends it seems to become rather hard to stay enthusiastic about that relation. Interaction, including discussing, experiencing and doing shared activities together is therefore an important part of social relations, especially friendship, but all social relations have a form of interaction.

A relation based on beauty is another aspect that requires more explanation. It seems so counter intuitive to argue that beauty is part of a social relation. Plato gives some clarity. He relates good to beautiful (216d) "The beautiful is a friend". An interesting claim, but what does it mean? Beauty is more than appearances, the beauty Plato refers to here, can also be in the form of virtues someone has. A friend that performs, according to oneself, morally right actions and is virtuous is quite important. Can you imagine being friends with someone who is morally bad, according to your own standards? Could one be friends with a person that he or she considers bad? Being beautiful in the eyes of others could be seen as a combination of attractiveness, goodness and intelligence (a 'beautiful mind'). Or in other words, a handsome, virtuous person. Naturally, the aesthetic part, attractive or handsome, is a property of which its inner workings are not so easily grasped. The study of such aesthetic beauty is not part of the subject of this thesis however.

2.3 A closer look at friendship

In the first half of this chapter we did not specifically focus on one type of relation over the other, although 'friendship' came up quite often, but without much explanation of what it exactly entails. In this second half, I will try to set-up a framework with friendship at its core that will help to understand social 'human-human' and 'human-technology' relations. In this half I will explain that the properties found in friendship can be used to understand all other relations, with the exception of that of the significant other. We will come back to that relation at the end of the chapter.

2.3.1 The properties of friendship

In the first half of this chapter several properties of human-human and human-technology relationships were introduced. Let us now discuss the properties of social relations. C.S. Lewis argued that friendship is something that should be shared, that affection and love for one's friend are important. Aristotle talked about motivations such as pleasure and utility, and most specifically the lack thereof. It is said that friendship requires both entities to be equal, independent of each other and morally good. Friends need to be virtuous, are beautiful. We have to admire the skills of our friends without enmity. We must see our friends as 'other selves' that one loves and cares for as if he is oneself. We must be empathic and we must proceed in shared activities.

We can bundle these properties into four categories. Affection, Utility, Interaction and Admiration. These four properties are not randomly chosen, nor were they chosen for their synergy with the thesis, but rather as a reasonable division of the properties mentioned above. These four properties are derived from the numerous properties we discussed. We could naturally also make a framework that treats love, friendship and affection as primary kinds of love -Much like Lewis (1960) does, so this framework is not the only way we could divide the properties. Affection and love however seem relatively closely related to eachother when such a concept is compared to, for example, shared activities. In order to maintain a reasonably simple framework, concepts that seem similar in nature were combined to create the four properties as mentioned above. The framework combines the simplicity of a few primary properties with the complexity of the different concepts that social relations have to offer.

This framework is useful because it allows us to compare facets of social relations to each other even though not all relations show all the same properties. Firstly the Affection category deals with all matters concerning love, affection and empathy. The Utility category concerns itself with things like pleasure, usefulness and the value of a social relation. The Interaction category will focus on shared activities, mutual interests and sharing friendships with others. Finally the Admiration category focuses on morality, independence, equality, virtue and admiration.

Affection	Utility	Interaction	Admiration
Love	Pleasure	Shared activities	Morally good
Care	Usefulness and value	Communication	Virtuous
Beauty		Ideas and discussion	Equality
Empathy		Mutual interests	Independance

2.3.1.1 Affection

What should we place under this category? Simply said, it is the feeling that you care for the wellbeing of a certain person. Affection is perhaps best understood as the things we care for. Caring in itself is a form of affection and empathy, as it being cared for. According to Lewis, affection is caused by likeness, more than anything else. And is therefore the most diffused and common of loves Lewis writes about, as there is no real competition for it, and it is quite easily attainable. A man can have affection for his family, friends and pretty much all his colleagues, neighbours etc. (Lewis, 1960).

It is useful to distinguish affection from eros here. The word love is used in various ways. One can say, "I love X" where the meaning of love is completely dependent on what X is. For example, if X is ice cream, the kind of love we get there is neither that of affection or eros, but rather a preference, a taste (quite literally as well) one finds enjoyable. If instead X is a person, we most likely have to do with affection. Loving a person generally implies caring for and being concerned for someone's well being. But this is not the only thing that could be meant with it. Next to the affection, eros can exist. Eros is the erotic, sexual love. The desire of another human. Although it is more than that, it is the feeling of being in love. This is more than just the desire for sex as (if made believe, in particularly by men) one can have this desire without being in love.

Lewis also classifies friendship as a form of love. He argues that friendship is the most unnatural of loves, there is nothing biological about it, as unlike eros or affection it does not incur a natural reaction (for example affection can cause jealousy according to him). It is also not productive for working together in large groups (as evolution might require of us humans to succeed), as a friendship with one of the others might make him or her more important than following orders of the leader. Which can indeed jeopardize an expedition. I think that the fact that Lewis categorises it as unnatural is already a sign that he might be wrong and it is not a form of love. Instead, I argue friendship is a combination of Affection, Utility, Interaction and Admiration. Affection, as discussed here, is our caring for others, in the way the 'love' friendship, in Lewis cares for friends. I will come back to the sharing argument of Lewis in the part about Interaction.

2.3.1.2 Utility

Utility is the part of a relation where its value and motivation shows. That friendship has value seems undisputed, although the nature of that value is often uncertain. Telfer argues that there are three values of friendship that can be found in Aristotle's work. The first being usefulness as a friend can help you in your time of need - While reversely

you help your friends in theirs. These can be considered among the duties one has toward his or her friends. The second value of friendship is that it is pleasant. It seems rather obvious that friendships bring pleasure, but it should be noted that friendship can also bring pain. Telfer (1970) argues that even though we can balance many pleasures and pains of friendship against each other (like the pain of losing a friend to the pleasure of gaining one), but that there are certain pleasures that have no mirroring pain. Notably the pleasures that come from shared activities, or as I put it, those that involve Interaction. The third value of friendship is that it is life-enhancing. It does so in various ways.

Firstly because it increases our involvement with the world, and therefore increases the number of things we care about, therefore feeling more alive. Secondly it does so because activities that are now shared (or perhaps the introduction of totally new activities like Cocking and Kennett (1998) argue) are undertaken with more pleasure than before. Finally friendship increases our knowledge, by giving us a different point of view to interpret the world. In Aristotle friends are primarily the ones that you do philosophy with, friends are the ones that challenge your world-views and argue for different points of view.

That said, we should not claim to say that someone is wrong for choosing to not have friends. As according to Telfer one is allowed to judge for himself whether or not he thinks he would be a good friend, or perhaps finds himself without time to involve him or herself with other people on a friendly basis.

Bacon (1871) also has an interesting conception of the value of friendship. He does not seem to try to understand what makes true friendship, but rather what benefits, or 'fruits', are the result of a friendship. Bacon distinguishes three benefits of friendship.

The first benefit Bacon describes is that friendship is an outlet for certain emotions in us. Most interestingly of those emotions is solitude, or feeling lonely. It seems rather obvious that being together with someone makes you less alone, but the approach Bacon takes here and uses this emotion as a reason for friendship is interesting. So far all the philosophers on friendship primarily associate friendship with the emotion of love, instead of seeing the fear for solitude as a cause and drive for connection. Although being an outlet for solitude is, according to Bacon, an advantage of a friendship, he does not claim that this is the only way one could combat solitude. Perhaps if we connect this to older philosophers, relations that are engaged upon so we do not feel alone, albeit without love, are those of common friendship that Montaigne talks about. Bacon concludes that this first benefit is primarily about sharing. Sharing 'doubles' the pleasure, and 'halves' the pain, he argues (Bacon, 1871). This claim of bacon is confirmed by some empirical studies (Pennebaker et al., 2001, Zech, 2000)

Bacon also argues that without being equal to one's friend, this sharing relation is impossible.

The second benefit Bacon explains is that of a truthful counsel without the fear of it being deceptive or incomplete. This is a valuable benefit because the insights and counseling friends provide might change one's understanding of situations, might help solve one's problems and helps to improve upon oneself.

The third benefit is that of help. There are many things one cannot do himself or by himself but would still like to see completed. This is where friends come in, as they are often willing to help. Bacon goes a step further and suggests that having a true friend is necessary for fulfilling one's role and having a good life. An interesting aspect Bacon describes here of the friendship relation is that it is a free relation. As Aristotle argued there are many types of relation, that of father to son, that of husband to wife, that of employee and boss. Aristotle suggests that also there is also a role from friend to friend. Bacon suggests this relation is roleless, free of any specific function.

All this talk of values might raise the question whether the speaking of utility and value is detrimental to friendship. Especially when Aristotle argued that true friendship is only possible if there is no motive for the friendship. Concepts like utility, pleasure or other motivations for such a relation seem misplaced. I have two responses to this apparent discrepancy.

Firstly we consider more relations than just true friendships with this framework. Social relations motivated by pleasure or utility might not be true friendship, but can be still a form of friendship. The difference between true and common friendships is according to Montaigne that those are based on services and benefits. In true friendship those are "not taken into account" (Montaigne, 1958).

Secondly the apparent lack of utility is a property in itself. One might call it melodramatic, but in this world we do almost nothing without a certain intention or selfish goal in mind. Egoism takes this stance to the extreme and argues that there is nothing we would do if not if it was somehow to pleasure or satisfy ourselves. Even helping an old lady across the street or showing someone the way would in the end be egoism, since we either expect similar actions in return (when we are lost or older for example) or do it to make oneself feel good because of doing a good deed. If we believe in egoism, real friendship seems impossible[2]. Altruism allows for relating to others for the sake of the other. Therefore it allows for the Aristotelian conception that true friendship is without a self-interest. Taking actions for the sake of someone other than oneself is quite special

[2]A way to work around this conclusion would be to see a friend as an other self, meaning that the ego would then include his or her well being and pleasure.

in itself, therefore the lack of motivation, or utility for oneself is also a property of a relation that is worth mentioning.

2.3.1.3 Interaction

Forming and keeping relationships of any kind involve a form of interaction. We could think of this interaction as being-in languages, verbal or written, and being-about subjects. Lewis argued that friendships always have an intention, something for it to be about. Therefore the being-about is what is most important for this category. Although having a common language is obviously also important for communication.

Interaction about specific subjects can be done in a variety of ways, including shared activities, talking about mutual interests. Also sharing the friendship among other friends can be seen as a form of interacting, as that will allow you to observe a friend in a way previously unknown to you (Lewis, 1960).

The name 'shared activities' is rather self-explanatory. Shared, meaning together, and activities, pretty much anything that both friends would like to do. Telfer argues that there are three types of shared activity, that of reciprocal services, mutual contact and joint pursuits. Telfer argues that we really need to have shared activities because just 'wanting' to act, is not enough. We should act upon our feelings (passions) otherwise there is no friendship.

In his lecture about friendship, Kant (1930) explains that it would be loathsome to share everything with a friend. He claims that there are things that for the sake of the friend, are not shared. Friends are not necessarily concerned with everyday trivialities that concerns the other. For example, one would not need to know whether or not his friend is married or has a sick father or any of those things. Lewis writes: "Do you see the same truth? That is the real question friends are concerned with".

As mentioned before, friendship is about something, as subject, that mutually interests both in the relation. This can be more than one thing of course. A very special kind of mutual interest would be in each others bodies, sexually. Lewis would put that desire under the love, Eros. Whereas in friendship the intentionality is towards an outside thing - In Eros it is pointed inwards, toward the relation itself. This also means that the interest in that relation cannot be shared with others. Lewis argues that unlike in Eros, in Friendship the number two is not the best as one can never fully bring out all the different aspects of a person on his own. Part of the aspects normally hidden from view for one friend, can come to light when the group of friends is bigger. "Dividing is not taking away, but rather increasing it for all".

2.3.1.4 Admiration

Telfer argues that there are three necessary conditions of friendship; shared activities, passions and acknowledgement. Shared activities and passions were mentioned in Interaction and Affection respectively, but acknowledgement finds its place in this category. With acknowledgement Telfer means to voice the objection that choice is required in mutual friendships. Acknowledgement functions as a way to check whether both parties have the same passions and as a social contract for further shared activities. Mutual acknowledgement of the friendship 'contract' is not all there is to this however.

Admiration, trust, respect, virtue and equality are all part of social relations. For example a difference between friends and family can be found in this category. The way, or reason, we respect our family is different than the way we respect our friends.

Virtue and what type of virtues seem to affect with type of love, and therefore the type of friendship that exists between persons. Aristotle writes that the love and friendship between father and son, is of a different kind than that between a man and a woman, or a ruler and its subjects (NE vii 1158b). Also these relations are asymmetrical in nature. As the love from father to son, is different from the love from son to father.

Equality in Aristotle's friendship theory is also a rather interesting concept. Aristotle differentiates between equality in justice and equality between friends. He writes that equality in friendships where one friend is better man (in status or virtue) than another, the better friend should also receive more love than the other. An increase in the gap between friends could cause a friendship to dissolve.

Cicero's conception of friendship seems in line with the role goodwill plays in Aristotle. Goodwill also seems to be a form of unconditional love rooted in human nature according to Cicero and Aristotle. Cicero goes further, and claims that being self-sufficient is required for entering into a true and lasting friendship, because when friendship is based on any expectations, fears or promises, a change of circumstances would most likely end it. (Cicero: D.A. 53) Cicero also tells us that self-confidence and being a virtuous person helps at making and keeping friends. Friendship is also based on the mutual acknowledgement of each other's virtues.

With this property I hope to capture the reasoning behind engaging into social relations, but also the reasoning behind keeping them. This seems to avoid the ancient discussion whether or not similarity plays a role in friendship. Empirical research done by Kandel and Denise suggest that similarity is relatively unimportant (Kandel, 1978). So here I argue that it is not similarity but instead admiration that plays this role. We do not require to have the same set of virtues, or the same set of skills or interest in order to be

friends. What is important in the friendship social relation is the acknowledgement of each other as one is. Other empirical research (GonzáLez et al., 2004) suggest that admiration and mutual respect are one of the most important characteristics for friendship in for example Cuba, whereas it suggested that in Canada shared activities are more important.

2.4 Applying the framework

Now that we have established our framework based on friendship theory, it would be nice to see how it works when we apply it to common relations. In the examples I will discuss a friendship relation, a family relation, a business relation, an admiration relation, and finally the relation to a significant other.

2.4.1 Friendship

If a framework based on friendship theory could not describe friendship properly, then there must be something wrong with the framework. Therefore the first step will be to show that friendship is indeed nicely described by the framework.

The shortest description of friendship using the framework would be: 'high affection', 'irrelevant utility', 'high interaction' and 'medium admiration'.

Let me explain. 'High affection' is the result of our concern for the wellbeing of our friends. Simply said, would you cry if person 'X' would die? The answer would correspond with the level of affection one would have towards person 'X'. The affection scale seems to be continuous, since it makes sense to ask whether you hold more affection towards person 'M' (your mother for example) more than person 'F' (i.e. your father). Even though both might be a relatively high amount, one could be higher than others.

Even though affection seems to be a continuous scale, I propose we denote affection as having four modes. High affection, medium affection and low (or No) affection and irrelevant.

'Irrelevant utility' comes partly from the point made by aristotle that true friends are not for the sake of any secondary reasons like pleasure or utility, but that does not mean that friendship is without its uses. On the contrary, friendship is quite useful for our well being, and much pleasure and utility can be derived from it (asking for a favour). Although also negative utility or pleasure can result of it, when friends ask for a favour, or the pain from the loss of friendship. Like affection, utility seems to be a continuous

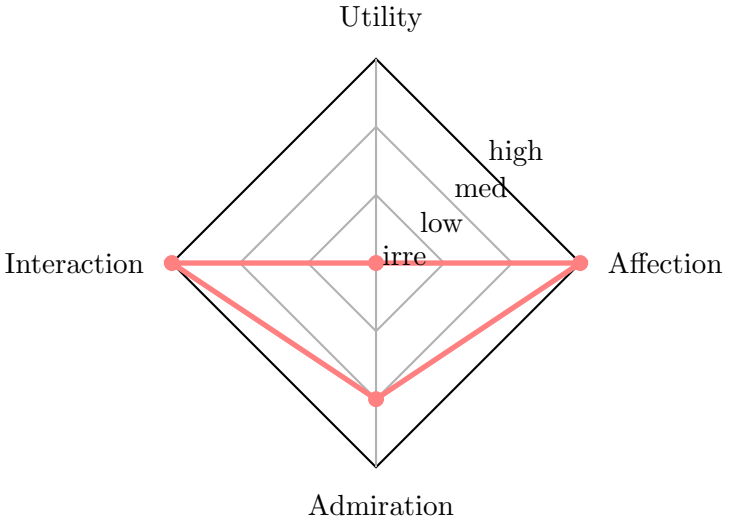


FIGURE 2.1: A friendship social relation.

scale, but for the purposes of the framework the four distinct modes seems more useful for comparison. Since for friendship the utility of friendship can vary greatly among the different friendships, and since it does not really matter, a value of 'irrelevant' seems most appropriate. One should not confuse irellevant with not required. It is quite required for a friendship relation to have an irrelevant utility, as it is one of its main characteristics.

'High interaction' comes from the points made by Telfer that we need shared activities to form and keep friendships. Furthermore we go out of our way to engage with friends in activities. Even though not for all friendships the number of times that one has shared activities is very high, the quality of the interactions are most likely quite high instead. Similar to affection and utility, interaction seems to have a continuous scale, and similar to affection and utility I propose the high-medium-low-irrelevant denotations for the scale so we can compare easily.

'Medium admiration' is the result of Aristotle's argumentation that we like our friends for their virtues, but that we posses similar virtues ourselves. We do respect each other, but it is not 'looking up' to nor 'looking down' on a friend, since a friend is truly our equal. Therefore it is 'medium admiration'. Similar to the other three properties I stick to the four-value scale even though the actual property seems to be a continuous scale as well.

One final point about this relation is that friendship is necessarily symmetrical, and therefore all parties in the friendship relation have these properties. We will see in the examples below that not all relations are of this nature.

I claim that the combination of 'high affection', 'irrelevant utility', 'high interaction' and 'medium admiration' symmetrically for both parties is unique to friendship. The challenge is to find a relation that has this same set of properties.

2.4.2 A family relation

The closest thing to friendship seems to be that of a family relationship, but as will be shown this relation still different in the way one describes it. Actually, family relations are quite diverse and have multiple instances and different interpretations. Note that it is definitely not impossible to be friends with one of your kin, and therefore have a relation that is even more like that of friendship than the relations I describe here. Also note that I talk about stereotypical family relations. I am aware that many families are broken and that these family relations will not apply to all families.

For this application of the framework, I categorise family relations in two groups. 'Parent-Child' and a 'Sibling relation'.

Let us start with parent-child, as that seems one of the more obvious examples. It seems clear that for both parent and child 'Affection' should be high. Since we can assume that a good parent cares for his or her child, and that a child cares for his or her parent.

The utility is different for both parties. Whereas the utility for the child is high, since he or she is quite dependant on the parent, the utility for the parent is rather low. Since generally speaking a (young) child is a burden. Of course pleasure and utility might result from having a child, but that should not be a primary concern. Therefore utility for parents is low or medium at best.

Interaction is high for both parties[3] as both engage in a lot of shared activities and conversation.

Finally admiration is also non-symmetrical between the relations, as a child should ideally look up to his or her parents. Whereas a parent should set an example for his or her child. Admiration for a child might come when he or she is grown up and is able of surpassing his or her parents on any ground. I will discuss such a situation later in the sibling relationship. Thus, the admiration for parent to child is low, whereas the admiration for child to parent should be high. Note that as the child grows up to be an equal, I argue that the relation becomes close to a sibling relationship and finally might

[3]It might seem that Interaction is necessarily symmetrical for both parties. As me having a high interaction with you must logically mean that you have a high interaction towards me. However imagine cases of stalkers or celebrities. In these cases there is often a one-way interaction channel that disrupts the symmetry of this property. One could argue that this is not interaction, a stalker could believe (wrongly) that a celebrity is in love with him or her and interacts with him or her through delusions of reference. See erotomania as a related psychological disorder.

even revert to a reversed parent-child relation when the child takes care of his or her elderly parents. Note that the relation is not necessarily restricted to child-parents, also grandparents, uncles and perhaps teachers can be described in this way.

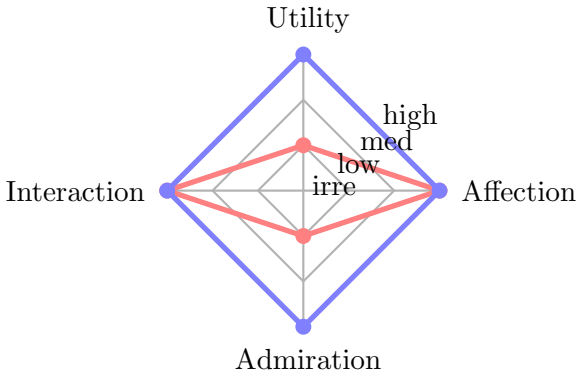


FIGURE 2.2: Parent-Child (red) and Child-Parent (blue)

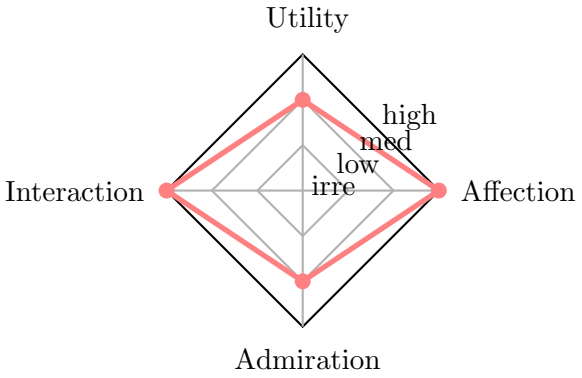


FIGURE 2.3: Sibling relation.

A sibling relation is a bit different than a child-parent relation. As for the most part siblings are more equal and often compete with each other at play. Depending on the age difference, this equality might differ and become more like a parent-child than a sibling relation. This equality can be found back if the properties are symmetrical between the siblings.

Generally speaking, a sibling relation includes affection not dissimilar to parent-child, so let us call that 'high'. Utility is a bit different. Both parties might help each other from time to time, but are far from dependent on it, and might even battle for the utility parents provide. In most sibling relations I would put utility on medium or low, but I will go with medium for now. We can argue that interaction is ideally high between siblings-relations, as they play and grow up together, but it is often the case that they have different friends-groups and interests thus therefore have less shared activities of quality than friends have. Admiration can be shifted a bit to the elder brother or sister, but is generally less than that to parents. Siblings should neither look up to, or down on each other (unless the age difference is high, in which case it is closer to a parent-child relation), thus admiration is medium. Note that the sibling relation is quite close to friendship, and that it is quite likely that we could categorise certain siblings as being friends as well.

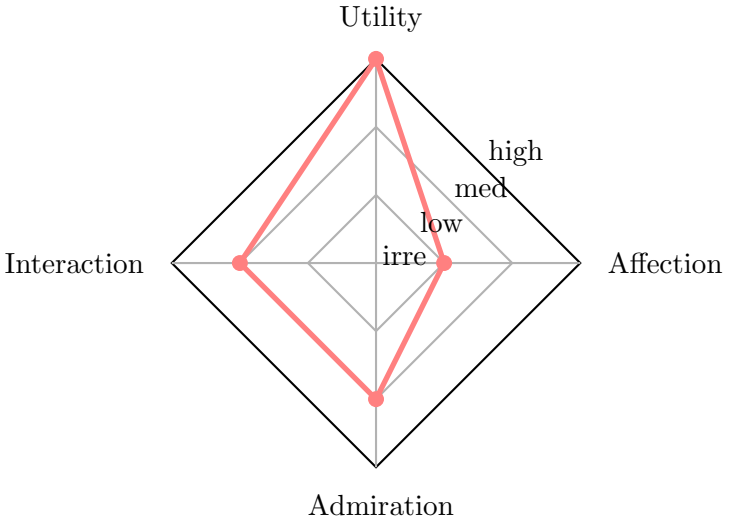


FIGURE 2.4: A business relation.

2.4.3 A business relation

I take a business relation as example, because it is a typical example of where utility is the most important attribute for the existence of the relation. Therefore I will argue that a typical profile of such relation is low affection, high utility, medium interaction, medium admiration. Furthermore, the relation is symmetrical with these properties.

Low affection can be explained by arguing that the essence of a business relation is indeed business. Affection has no place in the pure form of this relation. If a different party can provide the same service for a lower cost there is nothing that would stop me from engaging with that other party instead. Therefore affection between these two business partners is quite low. That said, it might be quite common for business partners to become friends, and therefore have affection for one another.

High utility should be clear, the only reason the relation exists is because of the utility it provides for both parties. If one of the parties can no longer fulfill the agreed upon utility, the relation is terminated.

Medium interaction needs some explanation. Business partners can often talk a lot, talk sometimes, or almost not talk at all. What is key for my choice for medium interaction is that the interaction is always about the utility in the business relation. In a sense it is the closest thing to a love relation, as it is the only relation to talk about the (conditions for the) relation itself.

Admiration is medium simply for the fact that business partners should not look up or down on each other. If you would not trust the other person to be worth their promises, then you would not engage into a business relation with that person.

Furthermore I argued that the relation is symmetrical. This is the case because all these properties should be the same for both parties. There is equality in a business relationship as it is based on mutual agreement, contracts and money. Both parties are dependant on each other. Also note that business relations can be extremely short in duration. In a sense when you walk into a store and try to buy something at the checkout, you engage into a short business relation with the cashier. Which is dissolved the moment you walk out of the store[4].

2.4.4 Admiration relation

You might have noticed that, so far most relations have had 'medium admiration'. So you might have started wondering if there would be relations that have admiration as their primary attribute. One of such relations would be a fan-artist relation. This relation is characterised by high affection, high utility, low interaction and High Admiration for the artist from the fan. While the artist has a low affection, high utility, low interaction and low admiration for the the fan.

High affection form the fan towards the artist can be explained by the amount of time a stereotypical fan would spend to find out all the news, songs or movies concerning that specific artist. I don't think a fan relation can be explained by just admiration for a specific artist, nor just affection for their works. A real fan should also be concerned with the well being of the artist, a high form of affection from a distance. For the artist it is low affection, because he probably did not know this specific fan existed.

Important to note is that the framework is designed for one on one relations, and not for relations regarding big groups of people. One could for example argue that even though the artist has little to no admiration for a single unknown fan, he does have an admiration for the fanbase as a group. In a sense this is not a direct problem for the framework, but it is not a subject required for this thesis.

High utility for the fan and for the artist. The fan gets pleasure from the works and/or existence of the artist, and the artist probably gets satisfaction or money from such fans. Low Interaction for both since they do not undertake any shared activities or interaction in general. The reasoning behind a low interaction (instead of irrelevant) is that irrelevant would allow for any level of interaction, while in this case it is clear that interaction is at a minimum.

[4] Although one could argue that in the case of a store, customer binding would require an ongoing social relation with a customer. This is not a social relation with the cashier, but instead with the company/brand.

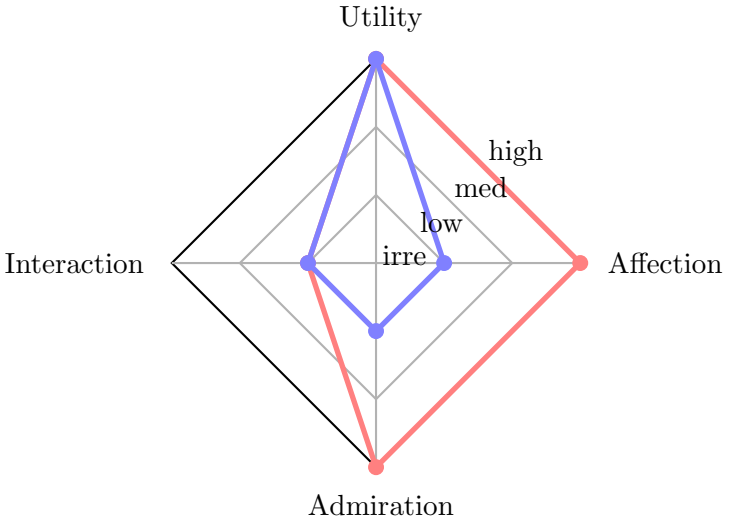


FIGURE 2.5: Artist-fan in blue, and fan-artist in red.

High admiration from the fan to the artist was the primary attribute for this relation. A fan really looks up to the artist in question. The artist however, as he is most likely not really aware of the existence of the fan, also has no admiration for him. Therefore I categorize that relation as low admiration.

Relations like stalker-victim or any of those sort are similar to this relation as it is similarly asymmetrical in terms of affection and admiration although that relation has more interaction.

2.4.5 Significant Other

Perhaps the most important relation for a human is that of the significant other, this is also one of the most problematic for the framework to handle. I argued before that the significant other relation does not really fit into this framework. This is because the relation is different in the sense that the relation is about the relation, instead about something outside of the relation. The only relation that has a similar property, the business relation, at least still has the business utility for the relation to be about. That said, lets see what happens if we apply the framework to this relation.

I think it goes without explaining that affection should be high in the significant other relation. Utility is harder, but I'd put it at irrelevant, since a significant other can cause pleasure as well as pain, the amount of utility really gained from a love relationship should be irrelevant unless some form of deception is in play. Interaction between significant others should be high, and admiration between the two should be medium since lovers, like friends, neither should look up or down on the other too much.

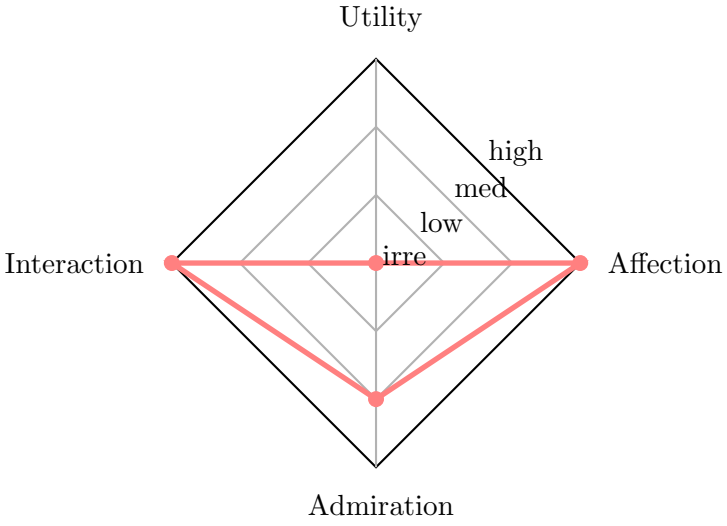


FIGURE 2.6: A significant other.

We can all imagine marriages that do not hold to these properties at all. A man might beat his wife, cheat on her and whatnot. We could even imagine marriages where all affection is gone. Perhaps from poetic considerations I'd like to reserve the term 'significant other' for relations that are working out. Relations in which both parties love each other. Let us move on with that in mind.

A close reader might have noticed that the values I gave to affection, utility, interaction and admiration in the significant other relation are the same as the values in a friendship relation. That would mean that according to the framework, there is no difference between a good friend and a significant other. This is problematic, as there is a clear difference. As Lewis argued, the type of love is of a different kind (as it includes eros) in a significant other relation. Also, as mentioned above, the intention, the relation is about the relation itself. Lovers talk of the love they have for one another. So if we wish to include significant others in the framework we would need to come up with another property.

This property could either be an aspect of affection, that denoted whether or not eros is in play. This could also help identify and describe cases such as prostitutes or to make it complete, 'friends with benefits'. I prefer however, to solve those cases by moving eros related activities to the pleasure utility.

Another option is to make a new property that denotes whether or not the intention (the reason for the relation) refers to itself or to an external object or subject. Internal relations, a relation that refers to itself, would be specific for love. Whereas the external predicate would denote all other relations.

I prefer the second solution as it describes best what the most important difference is and since the framework is meant to be a close description of reality it seems best to include it in this way if the addition of significant others in the framework is required.

2.5 Concluding

In this chapter I argued that we can make a framework based on the properties of friendship that can identify the different relations between humans. The framework has four properties; affection, utility, interaction and admiration which can each hold four values; low, medium, high and irrelevant. It also has two labels, symmetry and intention, which hold the values, symmetric/asymmetric and internal/external respectively. The idea is that each combination of those properties and labels denotes a unique human to human relation, and that it is exhaustive.

So now that we have the framework for human-human relations, can it be used on non-humans? The point of this thesis is to find out whether or not we can describe certain human-artificial agent relations in terms of human-human relations. This can only be done if it would make sense to apply the properties of human-human relations to those human-artificial agent relations.

In the third chapter, an overview of the technological possibilities of now and in the near future will be categorized into abilities that are needed to reach a certain value in one of the properties of the framework. The fourth chapter will discuss philosophical problems with assigning human emotions and thoughts to artificial agents. Finally in the fifth chapter we will discuss to which extent human-artificial agent relations can mimic or replace human-human relations.

Chapter 3

The capabilities of artificial agents in the context of human relations

3.1 Introduction

In the last chapter we discussed a framework consisting of four properties that describe important aspects of human-human relations. Before we can apply the framework to non-humans such as affectionate computers, we need to discuss two things. Firstly what entity is it that we are describing? And secondly, is it conceptually possible to apply these human properties to artificial agents? The former part of the question is what this chapter will be about. The second question will be dealt with in chapter 4.

So the concern of this chapter is finding out what the capabilities of artificial agents, robots or virtual avatars exactly are. One of the problems with ascribing properties like trust, virtue or emotions to technology is that these are too vague or too complex for developers to implement. Lucivero proposes a framework to tackle a similar problem. The problem she faces was the problem of reliable promises and expectations. One of the most important factors in this problem is that researchers and scientists are not politically neutral and must fight for funding. This means that they would often promise more radical achievements than is likely to result from the research. Lucivero et al. (2011) argues we should create a 'thick description' that looks at three things in order to achieve a more likely scenario. Firstly it should look at the technical feasibility of the proposed technology. This means that we should look at whether developers or scientists have a direction of a solution in mind when it comes to the technical challenges of the end-product. Secondly it should look at the social usability. Which

means that we should look at what actions will be required of humans in the various roles related to the technology. Finally we should look at the ethical desirability of the technology. This includes an effort to overcome the pre-existing ethical debate about the subject, since that might be based on wrong technical and social factors. Furthermore it should try to incorporate that our moral standards change in regard of new technological advancements.

So this chapter will use Lucivero's framework to look at the technical feasibility of robots and artificial agents of the highest level of complexity. Then we will take a look at the social usability of the robots and artificial agents. How would such robots be used? Finally we will take a short look at the ethical debate surrounding robots, as it addresses a core issue of this thesis.

3.2 Technical Feasibility

This section discusses the technical feasibility of robots and artificial agents. The first step for discussing the technical feasibility is to define exactly what is meant by robots and artificial agents, and sometimes affectionate computers or virtual agents. Note that this thesis is limited to the current and near future technologies, which means that a technology should not only theoretically be possible, but also in practice to some extent.

Since the goal of the thesis is to find out the highest type of social relation(s) (i.e. friendship) we could have with these technical agents there is no need to describe the lowest border of the technology, as it is of little importance to argue whether the DaVinci operating robot should really count as a robot or as merely a complex tool. Personally I opt for a strong definition of robots that includes a strong sense of autonomy and a complex artificial intelligence, but that does not solve the problem of defining robots. It only moves the border to a higher standard. Now the problem lies in defining 'strong' and 'complex' and show that certain things do not posses this property. An unrewarding task. Instead for any definition of robot and artificial agent one may chose the thesis will hold, as long as the chosen definition of robot is weaker than the one here. Weaker in this sense means that one allows strictly (thus only adding, not replacing) more to be artificial agents or robots than defined here. A stronger definition of robot than defined here might cause problems, as it requires more than a strong autonomy and complex AI, although there doesn't seem to be a realistic[1] stronger definition that would cause a problem.

[1] One could define a robot as a potato. Although stronger than my definition of a robot, that does not seem to be a realistic definition.

Wynsberghe (2012) lists a number of capabilities that robots, regardless of context, might have. Whether or not a robot really requires certain capabilities, for example the ability to move, is dependent on the context.

This list includes the ability to: Safely interact with humans, see, hear, move, feel and communicate.

3.2.1 Safely interact with humans

Being in the same space as a robot can be dangerous if the robot is not designed to cope with that. If for example a robot must touch a human, we must be sure that it does not apply so much pressure that it hurts or breaks human limbs.

How would we technically solve this problem? One way would be to not give the robot the ability to move at all, unless movement is strictly required for the context the robot is used in. If movement is required, we could prevent it from doing harm by making sure the exerted force is always lower than the force required to hurt a human permanently. This technical problem is by no means easy to solve, but all the technological means to achieve a safer robot exist.

3.2.2 See

Robot vision is a difficult ability that can be thought of in different ways. What we humans understand as seeing is different from perceiving, seeing also implies a form of understanding. For a robot, perceiving without further processing is about as useless. A robot has several ways to interpret what it is perceiving using its camera's or sensors. The first way, sensing, is that it could use that information to create a map of the environment it has to move in. If this environment is dynamic, for example because humans move, this map needs to adjust. For example this 'mapping' ability can be used to prevent collision with walls, furniture, other robots or humans.

A second way, recognizing, to imagine robot vision is that it can recognize certain objects for what they are. In the first way, recognizing a table need not to be different than recognizing a wall or any other object. In the first way robots sense collision hazards. The second way would entail a step more, recognize a wall as a wall, a table as a table and a human as a human, all with different properties. Naturally, this second way of interpreting is more complex and often unnecessary for simple robots such as a lawn mower robot. Recognition can have several purposes like differentiating between people using face-recognition.

Current robots, such as IRobot's "Roomba" vacuum cleaner robots show that the first way of seeing, sensing, is technically possible already. Most of these vacuum cleaner robots use infrared sensors or 'bumping' to sense their environment.

Recognizing is technologically not complete yet. Although companies like "Robot VISION Technologies" seem to make some progress in this area. RVT seems to be able to produce robot arms for the auto industry that can detect correct components and place them in the right location. Something that seems simpler than it is, since these components can be shifted and moved during transport.

The RobotVision@ImageCLEF 2012 challenge is a competition between universities that focusses on the recognition of rooms and objects. The top candidate of the challenge got about 80% of the possible 'points'. Which means that at that specific task it seems to work (Martínez-Gómez et al., 2012).

Reading, a specific skill of recognizing is easier than it initially seems. Many scanners are able to recognize letters and generate text files instead of images. Facial recognition is also in the works. Smeets et al. work on a 3D face recognition algorithm that can recognize faces regardless of their emotional expression - Which in turn should help with recognizing emotions in humans by computers as well (Smeets et al., 2012).

3.2.3 Hear

Sensors that register auditory information are easy to come by. Microphones can be found in many electronic devices today. Like robot vision, we can distinguish robot hearing into two ways of interpretation. Sensing and recognizing. A sound-sensor just waits till it senses a certain frequency, a set of frequencies or a pattern in the sound information. These can even be words. For example a robot that recognizes several voice commands does not necessarily understand the words being said. Imagine a copy machine that has two voice commands, print and copy for its respective actions, and imagine that a clever computer scientist pulls a practical joke and swaps the words and the actions around. Now giving the voice command copy, prints a document, while print starts a copy. This would confuse human users to no end, but the machine not at all. We humans have a connection between words and actions or objects, a discrepancy between these two leads to confusion. A robot with just sensing does not have this connection. This means that it does not mind, care, or understand that the difference between the command and the action is weird. For the robot, copy is print, and print is copy.

Like robot vision, the second way would be recognizing or interpreting. As mentioned before, recognizing is a rather different ability altogether, as it would require the mapping of words to certain objects or actions. For example by coupling it to a visual representation of an object. It would probably be best if this system is coupled to a reinforced learning system, that would allow the robot to construct it's own interpretation of certain words. An useful version of recognizing would be interpreting spoken human language. The ability to recognize, and understand a human sentence is rather difficult for computers. Also the fact that it is quite a challenge already to get the sound of a single person when surrounded by multiple sources of sound is not any easy task either.

The technical challenge of recognizing human language into text seems to be mostly solved now as most smartphones operating systems (Apple IOS, Microsoft Windows Phone, Google Android), Microsoft Windows and the Microsoft Kinect and many others already have a similar functionality, although it doesn't function on a human level. Dahl et al. (2012) try to improve upon current methods using a neural network with a learning algorithm. One of the problems they encounter in that system is that it takes a lot of computation time to train the systems. However, that means that as computers get faster and more complex (especially the GPU's) the practical ability of these neural networks should increase as it is more feasible to include more training data. It does not seem too optimistic to expect that we have capable (near) human-like speech recognition in the future.

3.2.4 Move

Moving or locomotion can be a difficult task. Moving independently even more so, as it will require a form of robot vision and a basic map of one's surroundings. Also not all terrain is equally forgiving. Stairs are an huge obstacle for the legless, human or not. That said, environments that are already build to accommodate wheel-based locomotion are relatively easy to navigate. Tests have been done by computer controlled cars that can sense the road and cars around them and have access to navigational information. These cars have driven around quite a bit without causing any accidents. A different form of movement would be the ability to move limbs or arms or so. These can be used to grab hold of things, point at objects, or in the case of the daVinci machine, operate on people.

The ability of movement in itself does not seem to me to be an obstacle for human-robot relations as that is primarily focussed on communication, perceiving and action. What could be important, in that regard, is the ability to communicate through movement.We will continue with that aspect in the subsection Communicate below.

3.2.5 Haptic feedback

Feeling is an ability that is rather useful when a robot has to manipulate objects or touch people. Note that the 'feeling' we are discussing here is not to be confused with emotions, but rather by our sense that allows tacitile feedback. Feeling would work by force sensors measuring the force applied to the robot. Another type of feeling could be measuring temperature where it touches. More complicated feel abilities would include the ability to recognize certain textures or map the feeling to an object. This kind of feeling is rather straightforward although it could be hard to implement in all its detail, there doesn't seem to be a reason that these technologies are in theory or practice impossible.

Another understanding of feeling does not only include physical pressure but also emotional feelings. These feelings would include difficult feelings like affection, love or a moral sense of right and wrong. Since we do not fully understand the ways these emotions influence our rationale ourself it is hard to replicate in an artificial agent.

Gadanho (1999) argues that there are several ways emotion influences reasoning that are useful for constructing artificial agents. These include; emotion as a source of motivation (Morignot and Hayes-Roth, 1994) and reinforced learning (Wright, 1996) , emotion to control attention (Sloman et al., 1994), a mood dependent memory system (El-Nasr et al., 1998), assistance in reasoning (Ventura et al., 1998) , behaviour tendencies in regard to stereotype situations (for example fear) (Botelho and Coelho, 1997) and the physiological preparation (arousal) of the body (Cañamero, 1997). These emotions only include emotions that are useful even when the agent is solitary as Gadanho explicitly leaves out the 'social emotions'.

3.2.6 Communicate

A very important ability for social robots is the ability to communicate. This ability can range from the simple way to utter some pre-programmed sentences to the ability to construct whole sentences from scratch. Language generating is very complex as a thorough understanding of the way language is constructed is needed. That would include the contextual double meaning of words, jokes, sarcasm or other inter punctual or contextual variations on the 'base language'. So far computers are only able to interpret languages that are not ambiguous. Every sentence, regardless of the context, should have one meaning. This is the way computer languages are build up, according to the syntax (grammar) a sentence can only mean one thing. Human languages are different, and are full of things computers may never understand (or interpret correctly), although they might come close.

Communication can be done in more ways than just spoken-language. Also body-language or emotions can play a role. Affectionate computing is a research area where work is done on computers that can recognize human emotions, and perhaps induce them through the expression using several bodily features, such as 'eyes', brows tilting the head etc.

Bates (1994) argues that robots can use artificial emotions to create the illusion that they are alive, and thus invoke a feeling of empathy in the humans they interact with. Breazeal (1998) argues that showing emotions can be used in learning robots to indicate whether the level of teaching is appropriate (bored or anxious). Klein et al. (2002) argue that a system that can recognize human emotions can help alleviate frustration or other emotions by expressing empathy.

Similar to Klein, Leite et al. (2008) argue that humans form relationships with robots easier if the robots behave in an empathic manner. Their empathic agent is said to be more trustworthy, caring and likable compared to a non-empathic agent.

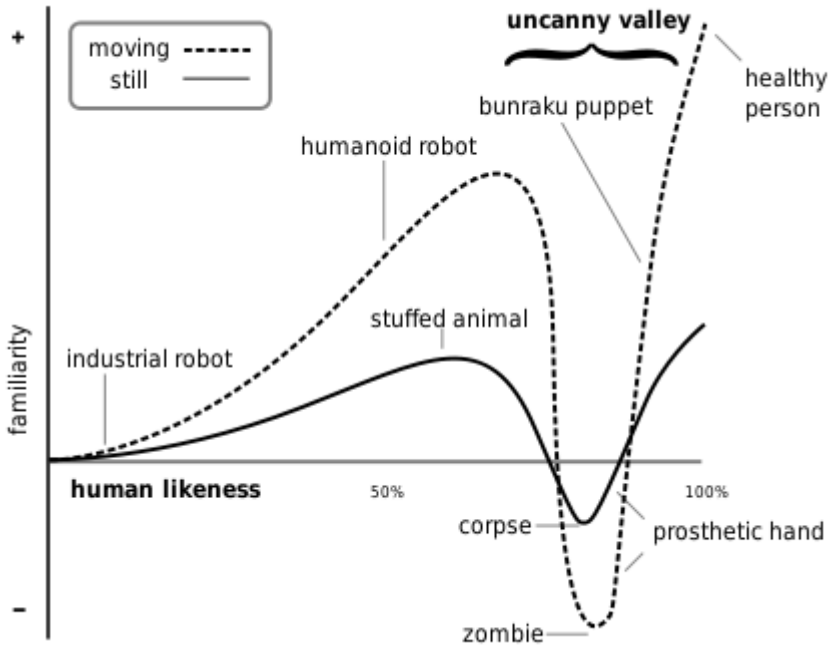
The robot in their example is the iCat, a cat like robot that utters sentences based on the moves played in a chess match. The choice for a game that has a clear evaluation function is evident, as it allows the robot to evaluate the board position and relate that to a player's state of mind.

This shows a problem for using this approach when designing general purpose robots, as those robots would need a multitude of evaluation functions depending on the context. A companion robot in that sense would need to be very versatile.

3.2.7 Appearance

Although hard to conceive as an ability, the appearance of a robot is quite important for several abilities. For example if it has human-like features it will be easier for us to recognize human emotions. Goetz et al. (2003) argue that is it important that robots appear and behave in accordance to their task. A playful robot (in appearance and behavior) is more successful in interaction with humans concerning a simple, non-serious activity often using a game. Whereas a more serious robot is better at communication when the task is of a more serious nature (a breathing exercise). The appearance of a robot also brings certain expectations of its capabilities. A robot that looks exactly like a human will always disappoint, as the appearance suggests it is capable of human actions, while it (for now) will always fall short. This is related to the uncanny valley effect.

The uncanny vally is a hypothesis about objects that the more an object looks or moves human-like the more these objects are perceived as unsetteling. Moving objects have this to a greater extent than still objects. For example corpses are less unsetteling or revulsing than zombies. There is a danger that androids (robots with human appearance) will be perceived as unsetteling when not moving the same as we do. Therefore it is easier for designers and developers of robots to largely avoid robots that look like humans for now (MacDorman and Ishiguro, 2006).



3.2.8 Non-feasible technologies

In the paragraphs above we can see that a lot of individual parts of an intelligent robot exist but still need some serious improvements until some abilities are at human level. However, there are also several abilities humans possess that are near-impossible to be successfully imitated by an artificial agent.

One that springs to mind immediately is that of emotions. As mentioned before Gadanho (1999) argues that imitating human emotions can be very useful for a robot and its artificial intelligence. However, what we cannot do is actually give a technology an emotion. At most a mental state that corresponds to a certain emotion and that compels the system to behave in a certain way.

Another technical problem is that of consciousness or being self-aware. Human consciousness is still problematic in cognitive science so it is not surprising that we would

have problems ascribing consciousness to a computer system, no matter how complex (Gazzaniga, 1997).

In short, the philosophical problem about consciousness can be seen as a result of Descartes' distinction between mind and body. Mind, consciousness and intelligence are three concepts that are often equated and all problematic in regard to artificial agents. The exact nuances and differences in the philosophy of mind debate are left for another time, but what we need to discuss in this chapter is how these properties could be ascribed to a machine.

Of those three, intelligence seems to be the easiest to ascribe. For intelligence is in the name Artificial Intelligence or Intelligent Agent. What makes those systems intelligent however? A chess computer that can beat even the best chess player is called an intelligent machine, but is it really? What it really does is compare a overwhelming amount of board positions with an evaluation function (that it did not construct itself) in virtue of specialized hardware built to do these evaluations faster and more precise than any human ever could. If you'd ask me, it is a pretty dumb system that overcomes its dumbness by using sheer calculative force. All its 'intelligence' it gets from how it appears to us, with a simple reasoning as follows; "The human chess champion is really smart, the computer beat it, thus the computer is smarter". So we are left with the problem of intelligence, can a computer ever be smart? Or can it only appear smart to us? Unless computers are going to write their own code[2], the intelligence of a system can only be accredited to its creators.

Consciousness and mind are similar apart from the influence of unconscious parts of the brain. A distinction that might hold in the differences between computers, since one could argue that a computer 'consciousness' might not be aware of all the code and programs behind its states. However this distinction is very useful for the purpose of the thesis, therefore we shall proceed to treat these two as the same in regard to artificial agents.

The first problem with consciousness in computer systems is that a computer (and humans for that matter) could never prove that they are conscious even if they indeed possess this ability. Imagine a conversation with a conscious robot and imagine what it would require to say to you so that it proves that it is actually conscious. It could claim: "I am conscious", but that is not sufficient. The only method for proving that a computer is conscious is that it can prove that it did not do what it was programmed to do. That it chose, from outside of its programmed code, to not execute a certain part of

[2] Which is definitely a possibility. Evolutionary programming involves the changing of parameters or code itself to optimize the result. As long as there is an evaluation or 'fitness' function at hand computers could potentially evolve their programming to achieve the highest score in the function. However finding a good fitness function is problematic for complex situations.

its own coding, for whatever reason. To me that seems like a logical impossibility, as an artificial agent will always be bound to its own code. This immediately shows that an artificial agent never possesses free will (as all its actions are governed by its code and input) and thus is not conscious. Unless it is possible to be conscious without having a free will (nor a free mind for that matter).

Several philosophers and scientists have brought forward arguments that it is impossible for any artificial agent to possess consciousness. Dennett (1994) replies to four of these arguments. These include counter arguments against that consciousness is something of the immaterial, is organic, or only present in born entities all that robots cannot possibly possess. His fourth counter argument is against that robots will never be complex enough to allow for consciousness. The problem according to Dennett with this argument is that if consciousness is merely a matter of complexity, then it is not so special after all, and thus no reason why it would be impossible in theory for it not to exist. This brings us to two possible solutions for the problem of consciousness. The first is that consciousness indeed comes from complexity, the natural constructivist approach. The second, deterministic approach, argues that consciousness does not really exist, even in humans, and is merely a result of our inability to ascribe the correct causes to our thoughts.

The natural constructivist approach (Gazzaniga, 1997) comes from the idea that as long as we can correctly mimic the functions and complexities of the brain consciousness will be an emergent property of the system. Naturally the problem of proving its existence remains, but we circumvented the problem that a computer would need to act outside of its programming. Rather its programming becomes so complex that it (or we) can no longer determine the exact causes of certain actions.

The deterministic approach takes this one step further. It labels consciousness not as an ability but rather our inability. The inability to determine the exact causes of our thoughts, choices and the inner workings of our brain. Robots might not share that 'defect' as they are more capable of 'knowing' the exact causes of every action. A solution in that sense is to program robots to forget or ignore this information.

That said, we must realize that all this is easier said than done, and that even the simplest 'consciousness' is so complex that it might "dwarf the entire scientific and engineering resources of the planet for millennia." - Daniel C. Dennett. (Dennett, 1994)

Searle (1982) has a different approach to why robots cannot possibly have a consciousness. He uses the famous 'Chinese Room Experiment' to show this position. In this thought experiment he suggests that he is locked in a room and has to communicate in written chinese. Assume he cannot understand chinese, written or spoken, and only has

a set of English rules that allow him to map a certain Chinese Input to another Chinese Output. His Chinese captors might think he does understand written Chinese, but in reality he does not at all. Programs necessarily do a similar thing, they map a certain input to output, and therefore can never have an understanding or consciousness.

3.3 Social Usability

Now that we discussed the technical feasibility of robots and artificial agents we must consider how this technology will be used in society. Obvious examples include care robots, companion robots, or military robots. We could also imagine many other uses for impressive artificial agents, for example as a personal assistant or as a companion in computer games, one that might even be considered a friend. Let us shortly describe scenarios in the field of care robots, the military, the personal assistant and that as a companion.

3.3.1 Care robots

Care robots are robots meant to perform certain caregiving tasks so that caregiving personnel are more efficient in their tasks. The incentive for these robots comes from a projected shortage of care-workers (Wynsberghe, 2012). It is suggested that robots will have to take over caring for the elderly to some degree. For the purposes of this thesis it is good to distinguish between two types of robots. Robots that help the care-personnel with tasks such as lifting, cleaning, or other physical tasks. And robots that are meant to become some form of robot companion. Current, simple examples of the second type would include Sony's AIBO and Paro, the robot seal.

Van Wynsberghe (Wynsberghe, 2012) argues that values are of importance in the design of a robot. Understanding the context and what is needed, is according to Van Wynsberghe, an important part of understanding the technological requirements of robots. According to Friedman (1997), values are what people consider important in life. He argues that each value is understood differently depending on the context. For example privacy can be understood rather differently by different persons.

So the values of the context in which a robot is to be designed are quite important. Brey (2010) argues that artifacts can promote or demote the realization of specific built-in values. Although this does not mean that these values are always actualized in the use of the artifact. Brey distinguishes between two types of consequences of built-in values. Expressive values and symbolic values.

Expressive values are values that a technology almost automatically endorses in virtue of its use, for example in healthcare the value of human dignity is important and realized through surgical tools by making them small, thus minimizing physical intrusion. For example this could mean minimise scarring. Thus, human dignity is a value embedded into the design of every surgical tool.

Symbolic values are more indirect, as they do not directly promote a value, but rather indirectly nudge people. For example products that claim they are green (energy wise) play on the values of sustainability, cleanness or goodness we associate with pure nature. Even though the technology might not have as a necessary consequence that those values are endorsed.

In healthcare, practitioners have claimed that two particular practices can help promote the value of human dignity. Namely touch and care. Both are meant as ways to keep in mind that what they are dealing with are actual humans, and that we should care for their wellbeing. Technologies that increase the distance, by reducing care and touch run the risk of objectifying the human being. Instead of a human in need of care the human becomes a defect object needing a fix.

Sparrow and Sparrow (2006) argue that robots will not be a sufficient replacement of human caregivers on the basis of their abilities, in the context of the aging human population. They fear that as humans get older, and less younger personnel is available we need to rely on robotic agents to help the elderly. Sparrow and Sparrow distinguish four ways in which robots are supposed to contribute to our wellbeing.

The first is that of robot 'butlers'. Robots that can help some elderly do tasks that would otherwise require the help of someone. Like help them get dressed, getting in and out of bed, opening doors, using other devices in house. Quite literally robots that can go fetch a cup of coffee or pick up a phone. Although these types of robots are complex in their ability to navigate houses, communicate with other devices and recognize human commands, these all seem hurdles that can be overcome. Sparrow and Sparrow note that we should be careful with these robots as the massiveness of these robots could cause accidents. There is a reason current industrial robots are kept far from human personnel. Although important, this is only a minor inconvenience. These types of robots need not be humanoid, and could be 'build into' the houses they occupy. Although this brings us to their point that it is probably much easier to do these same things without robots, but just smart environments and new technologies.

The second way robots can contribute is that of doing specific household tasks, like cleaning. Sparrow and Sparrow seem awfully conservative, as they bring up arguments like the complexity of the houses the elderly live in as problematic. Perhaps from a

technical point of view these might be obstacles that need to be overcome, but we could either change the houses to fit the robots or change the robots to fit the houses. Philosophically speaking, there are no problematic things this type of robot copes with.

A third is that of monitoring the elderly instead of helping them. The point here is that robots could act as a telepresence, doctors could help patients from a distance. These robots could make the elderly feel more at peace than if they would have some video camera in every room. Instead just one robot could suffice to monitor elderly patients when required.

Their fourth way is that of companionship. So called, social robots. Philosophically speaking, these are the most interesting for my thesis. These robots could be used to combat patient loneliness. Several toys, such as the Furby, Aibo or the NeCoRo could seen as simple social robots. A problem with these types of robots is that is it rather hard for them to be entertaining for a longer period of time. Sparrow and Sparrow argue that these types of robots are actually unethical, because they are deceiving people as they claim to be more than they are. We will come back to the ethical points in the ethical desirability section below.

3.3.2 Military robots

If movies like "Terminator", "The Matrix"," iRobot" or "2001: A Space Odyssey" are to be believed, we are doomed to war with robots we can no longer control. Thankfully, such robots and their reasoning capabilities are at least as far away as fully capable humanlike robots. Naturally the promise of warfare at a distance to minimize casualties is a promise militaries around the world value. We can get a good idea about which values are important for the military by looking at their current UAV and UGV programs.

The General Atomics MQ-1 Predator is a so called unmanned aerial vehicle (UAV) and is capable of delivering several explosive missiles to a target up to 700 km away. These robots have been deployed to war zones such as Iraq and Afghanistan in an increasing amount in the last decade (Singer, 2009). The Special Weapons Observation Reconnaissance Detection System (SWORDS) is a unmanned ground vehicle (UGV) designed to fight using many existing light and heavy machine guns, grenades or rocket launchers. Other systems deployed are robots to disarm bombs or scout for enemies, including one of the Roomba's (the cleaning robot) predecessors from the iRobot company.

What is interesting to note is that the robotic systems are now all remote controlled. Although some do offer some limited automation in the sense of automatic driving or backtracking. This lack of autonomy shows that one of the values for the design of these

military robots is control. 'Control' seems to be a general value in the military. Because of this control however, none of these military systems qualify as robots as defined in the beginning of this chapter, since they lack sufficient autonomy. Apart from being remote controlled there is little special about the systems involved from an artificial agent perspective.

Similar to care robots, these military robots display the value of safety, human dignity and power. Whereas in care robots the robots need to carefully handle the patients in a safe way, in the military the robots are the safety themselves. It is not the interaction with humans that requires the safety, but rather the presence of humans in dangerous situations. The robotic systems either prevent the danger or substitute for the human presence. We should note that human dignity (not having to do dangerous, dirty or dull tasks) is only preserved for the users of the robots. For the side fighting against the robotic systems a sense of dignity might decrease as it seems inherently unfair that a soldier can kill from a distance even though the 'victim' has no chance to do the same. Although it does not seem that we (generally speaking) are very concerned with this, as a similar argument can be made for every weapon invented, unless we wish to dispose of every weapon on these grounds. The value of power is in the ability of robots to increase reach, or the size of the fighting force without the use of more soldiers (Marchant et al., 2011).

Ethical considerations in regard to this technology should not focus on the artificial intelligence part of the debate but rather on the desirability of a new weapon. The ethical discussion seems to be similar to the ethical discussion: "is it right for a fully armored knight to kill soldiers wearing no protection?" - Assuming the armor gives a significant advantage. Although interesting, the current robotics do not change the question in a fundamental way. Therefore we will not discuss these military robots any further. Discussing the desirability of fully autonomous robotic killers starts being interesting when the military changes their values.

However if Marchant et al. (2011) are to be believed, actual autonomous military robots are on the horizon. This would significantly change the current way the military seems to approach robots and furthermore cause a wide range of ethical debates concerning lethal autonomous robots. However the robots Arkin (2008) and Marchant use might not fulfill the requirements set in this paper for autonomy in robots. For example, Tomahawk cruise missiles, mines and Phalanx missile defence might be seen as robots in regard to sensing their environment and acting upon it autonomously, but they do not significantly differ from ordinary weapons and still include a high level of control. Arkin argues that fully autonomous robots are an option for the battlefield and suggests a way to embed ethics into these autonomous robots. Although interesting in itself, it seems unlikely

that such systems are technologically feasible in the near future, since they will require a way to perceive the difference between soldiers, civilians and even soldiers without weapons. The embedded ethical evaluation causes the robots to become so much more complex that it seems unlikely that such system will ever be approved of in practice. Note that all previous 'robotic' systems that use lethal force are either controlled or incredibly simple and indiscriminate (being a bomb). Nevertheless, if advancements like these do come into the world it will definitely be a huge step for non-military robots as well, however these do not seem in the realm of possibility for the near future.

3.3.3 The personal assistant

The first intelligent agent with sufficient capabilities that will be the widely available will probably be an personal assistant artificial agent, because current mobile phone applications such as Apple's Siri and Google Now are backed by important companies and a large user base. The idea regarding robotic personal assistants is perhaps best illustrated by the Japanese animation series Chobits. In that fictional future world it is quite common to have small humanlike robots (persocoms) act quite similar to current smartphones. These robots include a good speech interface and a funny (customizable) personality. Although one could doubt the humanoid appearance will be there anytime soon, similar functionality can be put in smartphones. For example Modi et al. (2005) experimented with a learning artificial agent for agenda managing.

Perhaps the most used intelligent personal assistant application used now are Apple's Siri and Google's Google Now for their respective mobile operating systems. These systems function by accepting natural language as input and delegating input and location data to a webservice that then provides a result based on personal preferences. It can be used to ask for recommendations, directions, stock information, agenda management and many other things. Perhaps the most impressive part of Siri is its voice recognition and the parsing of natural language to interpret the actual request. The natural language interface is an important step to help bridge the gap between technology and users as there is less of a requirement for technical training (for example query languages) (Pizzocaro et al., 2013).

Furthermore, a natural language interface directly links to one of the attributes in the relational framework from Chapter 2, interaction. Although current speech recognition is still sub optimal, it is of paramount importance for interacting with robots in a, for humans, natural way.

The values that are important in this type of robot focus more on natural interaction, unobtrusiveness, and knowledge as a result of their functionality. Natural interaction

is an explanation for the natural language interface. Without this speech interface and another, for example, text interface we are pretty much left with a location aware search engine. Unobtrusiveness is a value that comes from the way users use these types of technologies. The only reason you'd want to be disturbed by your personal assistant is if you asked him/her/it to remind you of some action or appointment. At all other times it should be ready for use, but not in the way. Finally, the knowledge value comes from its intended functionality as a place to ask context aware questions. That would only work if it is generally capable to answer questions and requests more efficient than one could do it by him or herself. Otherwise users would not proceed to use the assistant. Therefore it needs to have access to sufficient information and combine it into the right request.

3.3.4 The companion robot

The social or companion robot is the most important candidate for the highest level of relation. In theory it could have everything, a speech interface not just specific to certain types of requests (as is the case with the personal assistant), but instead more general speech capabilities. Naturally the more contexts an interface needs to cope with, the more complex and difficult it becomes. Another advantage the social robot has over the personal assistant in terms of relationships is that it is by its very nature a less professional and more personal relationship.

One of the problems that is being tackled in the design of social robots is the question how to keep people engaged for longer periods of time. According to Gockley et al. (2005) the way to go is to give robots a personality, expressions and most important of all, an interactive and complex story. A robot should give the impression of having an interesting private life with goals, expectations, life changing decisions and perhaps even a love-life. Valerie, Gockley et al's receptionist robot, had such a background story and showed that prolonged social interaction with these robots is possible. The challenge for social robots at the moment is therefore not how we could give them emotions, consciousness or other attributes generally associated with human intelligence, but rather how can we make robots interesting, engaging and interactive for a long period of time. A goal that seems much more achievable for the near future.

This is where designers start to play a major role in the success of a social robot. Not only does functionality in the sense of voice interfaces and emphatic gestures play a role, also story, content and personality. Like Norman (2007) argued, there are several ways designers can affect the emotional bonding of users with inanimate objects. As argued

previously, (in Chapter 2) this is because humans anthropomorphize objects in order to understand them and automatically ascribe human emotions to them.

One of the ways designers can influence the users is by the functionality the object offers, but as we have seen, functionality is not all it takes when we talk about social robots. These social robots need to be unusual on their own. Successful objects are designed in a way that compels us to tell a story about their origins, a successful social robot will need to be able to tell us a story about its own origin. In that sense it will help a lot if a social robot can somehow remember and describe what has happened to it in the past. That way it also fulfills the function of a memory objects that can recite its tale from its own perspective.

Remembering one's own past is according to Aydin (2013) an important part of one's personal identity. Research that helps us form an image of what gives humans their identity can surely help when designing robots or artificial agents with their own personality. Although the problem with ascribing a 'self' to artificial agents is of the same degree as ascribing consciousness to them. Making an agent that 'knows' where its body begins and ends, is not enough for it to gain a concept of the self, nor it having the idea that it is embodied (Clark and Chalmers, 1998).

A second way (software) designers can greatly influence the success of a robot by making sure humans do not anthropomorphise with the robots in the wrong way, primarily by making them familiar. If for example the robot cannot remember previous encounters, we could think the robot does not care. Similarly if the robot seems overly attached to past encounters ("Oh so nice to see you, the last time was 141 days ago") as it feels alien.

If we are going for such a humanlike social interaction, its appearance should also be equally humanlike to prevent the 'uncanny valley' phenomenon mentioned before, but also to align its capabilities to its appearance. Not all social robots need to be this serious however, we can forgive lesser capabilities if the robot does not look like a human. For example, the relatively primitive toy Furby can be fun and engaging albeit for shorter periods of time. The most important values for social robots can be found in both the toys and the research projects. The values for social robots include, engaging, natural interaction, interesting, attentiveness and personal. Where personal both reflects on the relation between robot and human, but also the 'life' history the robot had.

3.4 Ethical Desirability

So now that we have talked about both technical feasibilities and social use abilities, we have an idea of what we should evaluate when we are talking about the ethical problems of robots. Ethical problems in Hollywood science fiction worlds (Matrix,Terminator etc.) where the means (to reach their goal) of robots has become to either use or dispose of humans, are quite irrelevant for now. This is because we have seen that robots are far from having an consciousness or a 'self' of their own and therefore will not be able to put themselves before us humans. What we should discuss however is less clear. In the sense of military robots, we could discuss whether or not using these devices to kill is ethical or not, but these ethical questions are not within the domain of this thesis. We could also discuss whether we would want care robots take care of our elderly, but there is no need to dwell on the socio-economic aspects of that problem, furthermore it is not really related to the thesis.

What we need to discuss however is the desirability of companion robots for everyday normal humans as an augmentation or replacement of their social interaction. Although the problem is situated in the future. Do would we want to allow that some social interaction is replaced by interaction robots or artificial agents? My intuition and probably that of many others, would say no, but why is that so?

An initial argument could be: "Spending time with other humans is superior to replacing social interaction with robots". Is this necessarily the case however? Why is that so? Is it because robots are not advanced enough to provide enough depth for a conversation? Is it because robots do not act sympathetic to our cause? These two arguments do not work, since technically there is no reason why future robots could not be better conversation partners than humans, or indeed do act sympathetic with us.

Another argument could be "It is not natural for humans to be social with artificial agents". While the statement seems to be true, these types of arguments are weak because they assume that what is natural is necessarily good. One could argue that it is not natural for humans to use computers and laptops, and here I am, writing this sentence on a laptop. Is it natural to replace social relations with artificial agents? Probably not, is that a problem? No.

The core of the problem is not that it is unnatural, but rather that artificial agents are per definition, artificial. In other words, none of their actions, feelings or interests are real, since it is all programmed in. One could argue that the fundamental difference between humans and artificial beings is that we have a free will. In the line of that argument this free will is what allows us to genuinely care and not just seem to care. Furthermore, that free will is essential for any form of equal relationship. Therefore the

best relation we could possibly hope to have with a robot is that of a robot slave and human master. In a sense robot users would only deceive themselves. Having a robot tell me I am a great philosopher, friend or gardener, does not make me great at any. Nor does it make me feel great if I know that I made the robot say that. A problem with these robots is that the makers try to make us believe that it is something it is not, namely an actual social being. If we would follow Hegel's master-slave dialectic (Cole, 2004) we would need to educate and free the slave in order to be able to acknowledge our own worth. The problem is that that would require genuine free will, emotions and consciousness in robots. Free will, emotions and consciousness are however not capabilities we can ascribe to current robots or those in the near future. If that would be the case, we would be having a whole different ethical discussion right now.

So, one could argue that a human-robot is necessarily not a genuine relationship but rather based on deceit caused by the artificial nature of robots. There are however several problems with this ethical argument as well. Firstly it assumes that free will is necessary for a genuine relationship. Secondly it assumes that artificial robots are indeed deceitful. And thirdly, it assumes that having a free will makes humans different than robots. These objections to the impossibility of any genuine relationship are further discussed in the next chapter, as there we will discuss whether these assumptions are wrong. If so, we will be forced to admit the possibility of genuine relations with future robots.

3.5 Summarising

We have seen that the artificial agents or robots discussed in this chapter are technically feasible, since they will not require genuine emotions or consciousness. However if a robot would wish to succeed it would still need to convince its human conversational partner that it does seem to have these capabilities. Whether or not that is possible will be discussed in Chapter 4.

Furthermore, from the use cases we saw that the robots with the most chance of having a strong social relation with humans will either be robots as conversational partners to combat loneliness in a care or social robot setting, or as an artificial personality on mobile phones or other devices. An advantage of these systems is that they will be more natural and engaging for their human users. Finally the ethical problems regarding care robots and military robots do not apply to the social robots as the goals and capabilities of these robots will be significantly different from the values the military or care institutions endorse.

Chapter 4

Appearances and reality

4.1 Introduction

In the second chapter, we discussed four properties; affection, usability, interaction and admiration, to describe the types of relations humans can be in. One of the important parts of these relations is that they can be asymmetrical. Which means that both parties in the relationship have a different set of properties in the relationship. This asymmetry is what can cause problems when we go to human-robot relations, since we humans might expect a relation to be symmetrical even though it is not.

This chapter focusses on the way we can deal with the gap between appearances and reality, as we discussed in chapter 3 that this gap necessarily exist. This chapter will present two different ways to view this gap and two different solutions as how to deal with it. These two views are about the step from reality to appearance.

The first view explains the difference using the concept of levels of abstraction. In short the idea is that there is a level of abstraction in which we could perceive robots as moral actors. We can use that same idea to think of robots as having emotions, a consciousness or other human properties that are otherwise not appropriate to apply to robots and artificial agents.

The second view argues for a different way to view the world altogether. In this view we argue that we should not just look at robots in a different way, but rather that we should look at ourselves in a different way. If we accept that everything in the universe is a function of cause and effect, then we can see that our brain, our consciousness and even our will are all caused by some process, making us not so different from machines indeed.

We need a final step in our argument. If you agree with either of the views above then the next step will allow us to use one of the views above and go from that appearance to reality. In the final part we will argue that appearances are sufficient, since we humans allow that appearances are sufficient when we deal with other humans. We do not question whether another human is conscious, even though there is no actual proof of that. If that is the case, according to this argument, why should we treat artificial agents differently? Why should we question their consciousness if we cannot even prove our own?

In response to this statement one could argue that we indeed might not be able to prove that other people are conscious, but that the chance that other humans are conscious is significantly higher than the chance that a robot is conscious simply because humans are significantly more complex. If this is an acceptable reasoning, accepting the rest of the chapter might not take much convincing. The points of view in the chapter show different approaches to which this same conclusion is the logical consequence. The conclusion that humans, apart from our complexity, do not have any special edge, soul or biological power that differs us from artificial agents.

The purpose of this chapter is to understand the way we think of robots and ourselves in order to appropriately apply the framework from chapter 2. As mentioned in the ethical debate in chapter 3, the three assumptions, 1. the requirement of free will for a genuine relationship, 2. the assumption robots are indeed deceiving humans and 3. the assumption that free will makes humans different from robots, need to be overcome in order to call any relation between humans and robots anything short of (self) deceit.

Before we start with the two views however, we discuss briefly the perception that being deceived by robots would not suffice. The point is that even if there is a bit of deceiving, it is not a problem for the appearances of robots.

4.2 The Experience Machine

A good way to illustrate why attaining a fake relationship or friendship is not good is the experience machine argument by Nozick (2012). This argument is meant to go against pure hedonism, in which a good life is measured by the amount of pleasure in ones life. One could argue that a part of the good life is in the communication with other humans with at its pinnacle the friendship relation.

The experience machine is a theoretical machine in which we somehow can fool the brain in a way that we alter our senses and memories in such a way to experience the perfect bliss at all times. According to hedonism, a good life strategy that seeks pleasure

and avoids pain, it is beyond doubt that we should plug into this machine, as it is the highest form of pleasure attainable at the lowest amount of pain. However, if we search our instincts most of us would not plug in. Why? Because at the moment one makes the choice to plug in, he or she is aware that it is all a machine and all the experiences obtained through the machine are fake. We want real experiences, real achievements, and also associate with real humans.

Similarly, the deceived businessman argument (Nagel, 1970) tries to show a similar point. This argument goes as follows;

A businessman is quite content with his life, he has good job out of which he gets satisfaction and his colleagues acknowledge and respect him. At home he has a loving wife and kids which whom he has a good relation. He is happy. Unknown to him however, his colleagues cannot stand him and make fun of him when he is not around, his wife does not love him and cheats on him, while his kid despises him. They are quite good deceivers so they can all keep this up indefinitely. Can we really say he has a good life? As you can see, this argument is here to show that our perception of a person's happiness quite depends on what we know of our or another's situation.

One can wonder, does the deceived businessman have a good life? If you think yes, then this chapter does not hold many controversial things for you. If not, then the following arguments might convince you otherwise.

Let us try this by reversing the question. Why would he not have a good life? Perhaps you think, because his happiness is built on lies. While true, this is not necessarily a problem, the man himself does not know it, and will not find out in this hypothetical situation. If having any lie in your life would make your life unhappy per definition we should immediately stop making children's life worse by allowing Santa Claus to exist. There are enough examples to show that depicting the world untruthfully can indeed improve life in certain occasions. Formally we need to distinguish internal mental states from external state of affairs. What we argue here is that the as long as the internal mental states are at least equal or higher than the external state of affairs allows for (assuming those are good enough), then in the long run it is a good life even if it requires being deceived about the external state of affairs to reach that level. The problem is that our internal mental state as external observers would not allow for the same happiness attained by the deceived businessman because we are aware that it is way above the happiness level the external states of affair would allow for. Simply because a large gap brings a large risk of a miserable life with it.

Feldman (2002) argues for a similar solution. However he would classify the above solution as biting the bullet. Since we basically argue that the life of the deceived

businessman is not as bad as it seems to us. Feldman however also has a way to arm hedonism against this situation by stating that pleasure derived from true states of affairs are more valuable than pleasures derived from untrue states of affairs. While this is a useful addition, it seems unpractical to me. Even though having a way to externally verify whether someone ought to be happy or not is useful, what really matters is whether a person actually is happy. To compensate for the risk of being deceived we ought to use a function including the length of time, intensity of the pleasure, truthfulness of the states of affairs, the risk of discovery of these true states of affairs and the impact of that discovery. In the case of the deceived businessman the impact of the discovery would be incredible, so much even that the slightest chance of discovery would be unacceptable. This way it is also no problem to discover small lies or wrong perceptions that have little consequence (for happiness) that exist everywhere in our world. As would be also the case of friendship with robots.

4.3 Level of Abstraction

Before robots can actually deceive humans to have a consciousness, we need to solve the problem of how an artificial agents appear to have a consciousness? As was argued in Chapter 3, it is really hard to prove that an agent or human is conscious. What is easier to perceive however is whether an agent makes similar and logical choices. For example it should be able to interact with its environment and adapt according to the situation, furthermore it should be autonomous. An artificial agent should be a moral agent on a certain level of abstraction.

Floridi and Sanders (2004) argue exactly that with their Level of Abstraction (LoA) argument. Which could be summarized as follows: something is an artificial agent on a certain abstraction level if on that level of abstraction it can be perceived as: Interactive, Autonomous and Adaptable.

This level of abstraction argument is useful when combined with the ability to make moral choices. However, for a social relation just the ability to make moral choices is not sufficient, since it also requires a form of affection, interaction, admiration and usability, as argued in my social relation framework in chapter 1. One could however argue that for a good moral decision one would need a proper social relation to the entity the choice is about. However this is not the case since we humans make many passive moral decisions about people we barely know.

In Floridi and Sanders the properties, interactive, adaptable and autonomous can be that on any Level of Abstraction (LoA). While it is important that it has to be an

artificia agentl no matter what the LoA, because it needs to have these properties on lower levels to be able to have them on higher levels. Let me clarify this with an example Floridi (2008) uses to explain LoA, the game of chess.

The board of chess could be represented as a two dimensional space of eight by eight squares. This representation in itself is a level of abstraction, we could argue about how high this abstraction is, since we could also describe the board on a really low LoA as atoms moving around, which is something we will come back to later.

Moving to a higher level of abstraction means that there is more abstraction and therefore less micro actions observable, which means that more macro-level actions become apparent. Moving to a lower level of abstraction does exactly the opposite, where more smaller parts or processes are visible and those somewhat obscure the higher level processes.

Let us assume the average level of abstraction where we represent the board in two dimensions, where the columns (or in Floridi's words Files) have character indications and the rows (or Ranks) have number indications. All the pieces have also their own characters. This is the level of abstraction chess players would use to write down games on paper. For example a move with a rook would be; R a1-a8. Even if we knew nothing about chess, we could get a full understanding about the game just from observing enough games by deriving all the rules. If we now go one abstraction level higher, as Floridi does himself, to the Row Chess or the Column Chess we can see that we lost a certain amount of information, as we can no longer see position of pieces in columns or rows. For example in Column Chess the movement of pawns would be impossible to detect apart from their ability to move columns by removing a piece of the enemy in an adjacent column. If we were to solely observe chess from this level of abstraction we can only find a limited number of rules. A limited number of abilities of the game. A lower level of abstraction always has more. Furthermore a lower level of abstraction always contains the abilities the higher level of abstraction has. We will illustrate that by bringing the chess game to an even higher level of abstraction. In our previous abstraction, we brought chess from a two dimensional game, to a one dimensional game, a game of column chess or row chess. If we go one step further, a zero dimensional game of chess can be observed. In this dimension there are neither rows or columns, thus all the pieces are just there. The only things we can observe in this game of chess is that there is a certain amount of pieces, sometimes pieces are removed from the board. The king-piece never gets removed and that pawns sometimes change into different pieces. It would be impossible to get a full understanding of the game with this zero dimensional chess, but all the rules that we can infer, can also be inferred in Column chess and Row

chess individually. For example it is obvious that the number of pieces one starts with and that sometimes pieces get removed can both be inferred from Column chess.

Also that the king-piece never gets removed is pretty obvious, and yes also the final rule that pawns sometimes change into a different piece can be observed. So going down levels of abstraction always retains the abilities it had on a higher level, but higher levels of abstractions can abstract away certain properties. The rules are still there, we just cannot observe them. Similarly if we move from two-dimensional chess to three-dimensional chess, the pieces suddenly get a shape, weight, height, therefore increasing the complexity of the situation. Still all the rules that could be observed in two-dimensional chess are applicable. Even if we move to a level of abstraction where we can observe the individual molecules of all the pieces in the board and their respective positions, the rules of two dimensional chess hold. Note that at the level of atoms, where one describes the movements of all the molecules around the place where the game of chess is taking place it becomes extremely difficult to keep observing the relatively high abstraction process of the game of chess.

One could therefore argue that the game of chess is an emerging property of a higher level of abstraction. This chess as emerging property is similar to consciousness would be for our brain. It is very hard (perhaps impossible) to observe something like consciousness on the neural level, but it is rather easy for ourselves on a much higher level of abstraction. If we cannot observe a process on a low level of abstraction, but we can on a high level of abstraction this can mean two things. We either do not understand the low level of abstraction significantly enough to observe the effects for the higher level process, or the high level process is a by-product, an illusion so to say, of the lower level processes. If you wish to believe in consciousness and free will, one should adopt the first point of view regarding the workings of our brains, otherwise the second view will be more suited.

So what use is this clarification of level of abstraction? If agents have adaptability on a certain level of abstraction, they necessarily also have this property on lower levels of abstraction. Therefore an artificial agent does not require a set level of abstraction, the notion that it has the properties on any level of abstraction is enough to ensure it has it on sufficiently low levels as well. Surely we can always abstract so far away that we are unable to observe any of these properties, but we can do the same for the human brain. Floridi and Sanders argue that -not- knowing the source code can make a program seem as if it is adaptive and interactive on a certain level of abstraction, therefore they call it an Artificial Agent on that LoA. One could argue that a weakness in this argument is that for the programmer, who knows the source code, it is not an artificial agent, as it merely follows the instructions he wrote down. There is nothing

autonomous about the agent. If you agree with the argument about chess, then if an artificial agent is autonomous on the LoA where one does not know the source code, it must necessarily also be autonomous on a lower LoA where one does know. This results in a contradiction. Furthermore, it is also autonomous on higher LoAs; it might just be unobservable on that level. The consequences of making the LoA irrelevant is that artificial agents suddenly require a much more complex system, because they must also have traces of the properties of autonomy, interaction and adaptability on very low LoAs. We can no longer fool ignorant users; just seeming to have autonomy is no longer enough. It needs to have it.

One could argue that higher levels could produce 'illusions' that do not exist on lower levels, but are definitely needed to explain the situation on a higher level - Consciousness as will will see below is one of these 'illusions' one could argue. However, we must be careful not to mistake these 'illusions' for real things, even if some of these emerging illusions is very useful to explain certain things - For example, chemistry or biology compared to physical processes. However useful (and perhaps detectable) these representations may be, these representations are not necessarily reality if there is no similar effect on a lower level of abstraction. Chemical and biological processes exist on a physical level of atoms, although they are harder to explain. It cannot be that chemistry tells us some molecule has become a hydrogen molecule while physics tells us otherwise. This is the way we should treat these emerging properties on higher levels of abstraction. If there is really something that, physically speaking, seems more than its parts it is either an illusion or magic caused by our lack of understanding of the lower levels of abstraction.

What about humans? Do we have autonomy on low levels of abstraction? Yes, we do. If we look at the low LoA of brain functioning, even at that level we can say that we are interactive, adaptable and autonomous. Even if we knew exactly how the brain works, it is still the place decisions are made and input signals are processed. Unless one wants to go the determinist way, who would argue that there is neither free will nor autonomy in that sense for humans. Which is a fine point of view as well, as it just shows that we cannot use LoA to differentiate between humans and machines. Thus a non-determinist would say that humans are autonomous on even the lowest level of abstraction, and so are artificial agents. While the determinist would say that humans nor artificial agents are autonomous on any LoA. Either way the ground is paved to stop differentiating humans and robots merely on their biological properties. Robots with the required properties could be just as responsible for their actions as humans are. Since robots are at least as capable to reason properly there is nothing conceptually wrong anymore about assigning responsibility to robots. It was wrong in the eyes of Floridi and Sanders because assigning responsibility to an agent that does not have autonomy on

a low LoA would not make sense (which is why they introduced moral accountability), but now that it is required to have autonomy on every LoA, this is no longer a problem.

My criticism on the LoA argument makes it significantly harder for robots and artificial agents to become advanced enough and have the necessary properties to become fully capable of social relations. Robots need to become more complex to such extent that we need to discount the probability of it happening anytime soon. We will have to work with what we have, which means simpler agents, simpler programming and simple solutions to the problem of creating an agent that appears to be conscious. The key is to look at the reasoning behind the introduction of moral accountability that Floridi and Sanders introduced. If we can create agents that are according to them morally accountable but not morally responsible then we could perhaps also create agents that are appearing conscious instead of being conscious. The LoA argument shows us how to proceed: we need to hide the deterministic nature of the programming code. The users that will engage in the social relation will have to observe the agent on a higher level of abstraction. A level where the code is abstracted away, but the semi-intelligent behavior is still observable. Appearing moral seems to relate to morally accountable in the sense that something that appears to be moral but is not in reality, is still morally accountable since it does behave as a moral agent. A being or agent that not only appears moral, but also is moral, is not only morally accountable but also morally responsible for its actions. This means that artificial agents that are morally accountable, according to Floridi and Sanders, also appear to be moral agents.

Interestingly the level of abstraction regarding appearances is in reverse order to the levels of abstraction regarding being (i.e. whether something is instead of seems to be). The higher the level of abstraction, the more it must rely on appearances and the less on being, while the lower the level of abstraction, more of the being is important. Even a colony of ants or bees can appear to be conscious if we move the level of abstraction to a level where we cannot observe every single entity.

This is good news for the social robot enthusiasts, but it begs the question of which level of abstraction is required for current robots to appear conscious. Furthermore what is really important is whether the level of abstraction that is attainable through that way is actually lower than the level on which we observe the robot. However, specifying exactly at which level of abstraction current robots are conscious is a task beyond me. What can be done however is specify the highest level at which it should appear conscious in order for the social robots to have a real shot at forming a genuine social relation. This level is the level at which the users observe the robot. It is key to keep (thus not to lower the level by adding knowledge about the construction and programming of a robot, similarly, questions about the consciousness of humans only make sense when we

investigate the workings of our brain) the level on which we observe the robot as high as possible, so the level would be similar to the level we observe other humans at. The LoA where none of its individual parts are clear, a level where the robot appears to be a body. Its conscious appearance should be in accordance to its body. On a lower level we cannot really speak of it having a body, but that is not so important. What is important is that it is clear that the robot does appear to have a body and consciousness on the relevant level of abstraction.

The gap between appearances and reality can be seen as level difference between where the appearances stop and the being begins. If the being and the appearance overlap, then there is no gap. If the being and the appearances are significantly different (for example a computer that appears conscious) the gap is huge. Especially if the illusion of appearance falls apart if one moves a small step down in the levels of abstraction. To reduce the gap we could either alter its being in a way that corresponds closer to its appearance, or we could try to improve the illusion by bringing the appearance down a few levels in abstraction. By using neural networks, evolving functions or other ways to obscure the inner workings of the agents we could lower the level of abstraction where the appearance still holds. Changing reality (electrons in copper circuits) to something more organic (not simply because it is organic, but because of its more complex functionality) or perhaps using quantum computing could bring the reality closer to the appearance. Either way would improve the appearance of consciousness in robots and artificial agents.

Thus far we have tried to make robots more like us, because we anthropomorphize robots and objects to suit our observations. That seems to come easy to us, but concepts like consciousness, free will and the mind are troublesome to apply to machines to say the least. So instead of trying to reproduce these in robots we could try a different approach. An approach where we would have to let go of our problematic view of the self and see ourselves as a really complex robot. This view is what is being discussed in the next section.

4.4 Determinism

Whenever someone utters the word determinism it is not unlikely that a person in the room has serious reservations about determinism (unless it is a room full of computer scientists perhaps) because it is being associated with concepts like fate and predestination. Although this association is not entirely incorrect, it is not what the focus of determinism should be. Furthermore, we should not underestimate the illusion of free will. First we discuss a brief overview of scientific determinism and put some scientific

arguments against determinism. Then determinism is used to explain a point of view in which we humans are not so different from robots after all.

Scientific (or causal) determinism is an idea that follows from the scientific revelations that natural phenomena obey definite scientific laws (Hawking, 2013). Laplace suggested that if we knew the location and speed of all the particles in the universe we could calculate the past and the futures states of the universe, in Laplace's (Weber, 2001) thought experiment Laplace's demon has all this knowledge. In view of scientific determinism the whole universe could be seen as a single line (timeline) of states of which the first state will necessarily cause the final state to happen. There are no external forces (forces external to the universe) that could change this path.

According to Hawking, there are two problems with applying scientific determinism. The first being that quantum mechanics shows that we cannot know both speed and the location of a particle. Since when we measure one of these properties, we influence the other. This principle is called the uncertainty principle. The uncertainty principle therefore implies that Laplace's demon could never be simulated.

A second finding that undermines the required knowledge we would need for Laplace's demon is the existence of black holes. A black hole would eat information, light that goes in never comes out, the information about the particles there can never be known to us or any observer. Thus every black hole is a source of a fundamental lack of understanding about the particles in our universe. Furthermore, the existence of numerous micro black holes would also obscure information even when no apparent (or big) black holes are around.

Interestingly both these objections seem to criticise Laplace's demon on its knowledge requirement, not on its implications. The first is a criticism on the precondition of Laplace's demon, namely that one should know the speed and location of all particles, thus since obtaining both is impossible the demon is impossible. The second criticism is slightly different, namely that even if the first criticism does not apply and we really could have all information, the loss of information while calculating the future would bring further calculation to a halt.

That it is theoretically impossible for anything to know both place and time of every particle in the universe is indeed a problem for the existence of this demon and his ability to see the future. However the idea behind the demon still stands. The point of Laplace's demon is not to show that we could know the future if we had enough information, but the underlying principle that the future follows from the present. Both criticisms of Hawking undermine the knowledge requirement, but not causal links the idea stands for. Therefore the general rule of determinism, the future follows directly

from the past, still stands. Said like this it seems quite obvious, the future follows the past. Who would not agree? However you need to realize what this entails. We humans are also a part of this universe, all living things are, and they adhere these same rules. Your future self is a necessary consequence of the present self. Your future self in 1 millionth of a second is a necessary consequence of the current self, which in a sense means that from the start of the universe till now nothing free of the rules of causation has happened. This includes free will or choices in general, they simply do not exist in the sense we are used to. We should instead speak of decisions and will (not free) or goal. In contrary to choices, decisions do not imply a seemingly baseless course of action. Instead decisions are directed, have an intention, towards a goal or will. Also this goal or will is not without any grounding, there are reasons (perhaps unknown to oneself) that guide and influence every will, goal or decision.

The step from talking abstractly about the particles in the universe to our behaviour seems like an extremely big one. The reason this form of scientific determinism can be useful when thinking about humans is that also our brain, our body and our environment are all obeying these same rules. Therefore even for us, all our actions, decisions and idea's are causally linked to previous configurations of our environment. So without delving too much into the pitfall that is philosophy of mind, one could argue that on the basis of scientific thinking we should accept that we are in fact determined by everything around us and our past.

The next step is even more important, as we will have to compare humans to machines, and look for fundamental differences. If we really do not have a free will as determinism suggests, but rather all our decisions are somehow grounded in reasons, how are our brains then different from an extremely complex decision tree? In that sense, software is our best way to construct complex decision trees so artificial agents and robots are our best shot at creating something similar to us as soon as our models for human behavior are advanced enough and go beyond the simplest of contexts.

There are however several arguments that could be brought up against the premise that humans are essentially no different from robots. One could argue that humans are different from machines since humans are biological, and machines are not. However, just the organic nature of human beings is a weak argument. It would have to be deduced to that we cannot hope to mechanically imitate organic structures, perhaps because these organic structures are too complex. There are two answers for this objection. Firstly, we could use simpler solutions to achieve similar functionality as the biological counterparts, take for example artificial hearts, eyes or ears. Yes they are not as advanced as a real heart, eye or ear, but they can function as a replacement. Perhaps we cannot hope to reproduce all the things in an organic creature, but neither do we have to. What

biological function makes us human? There doesn't seem to be that one thing we cannot replace apart from the brain for now. The most interesting claim about that is about consciousness, a subject we discussed more in Chapter 3. The second objection comes from the idea to use organic parts in computers. If we really cannot mimic certain organic functions (because of their efficiency or complexity), then we could employ these for simple functions.

This point of view allows us to view the gap between humans and robots not to be based on concepts like consciousness or free will, but rather on its abilities compared to our own. The gap would be how it is lacking or perhaps even superior to us in certain areas. For example physical strength is a property that robots can very well exceed us in already. While on the aspects of context awareness, adaptability, speech generation and recognition robots are far inferior to humans. This approach shows us a bottoms-up view of what we should start improving in order for the robots to be more humanlike.

Similarly to Level of Abstraction approach however we can never prove a robot to be conscious, have a morality or any of such things. Making a robot perfectly moral might even be purposely different from how (many) humans are. A morality totally based on logic might seem inhuman or not caring. Such endeavor might give some insights in the workings of our own moral reasoning. One's moral intuition might vary to be depending who you speak to, which would make the perfect human-moral robot even more controversial. In any case, the best we can prove is that we could perhaps make robots that appear to be moral and conscious. Which in the case of determinism is similar as what we humans do ourselves.

Thus far we discussed two ways to view how robots could indeed gain the appearance of morality or consciousness. In the next section we will discuss that these appearances are enough to base genuine relations on.

4.5 Moral Appearances

One of the main goals of this chapter is to discuss the question of whether a relation based on appearances is enough. In order to appropriate the importance of appearances we discuss several philosophers or scientists that used a similar approach.

4.5.1 Interrogation games

In 1950 Alan Turing (1950) faced a similar problem with artificial agents as this thesis. He faced the question, "Can machines think?". As we have discovered this is a troublesome question. Turing's solution was to use an imitation game. In this game there are a man, a woman and an interrogator. The goal for the interrogator is to guess who is the man and who is the woman. The interrogator can ask questions to one of them and that one must answer. The goal of one of the players is to fool the interrogator and to make him think that he or she is of the opposite sex. The final player's goal is to help to interrogator come to the right conclusion. Naturally the interrogator is situated in another room and the questions are answered by writing (through a computer or typewriter preferably). Turing's question now became, can a computer fool an interrogator as many times as the human players?

Turing briefly discusses several arguments against the validity of this test. Interestingly some are quite similar to the arguments we already discussed before, it is surprising how little progress there has been on this subject in the past 60 years. However, since there are also a few new arguments we will take the time to briefly go through these objections. The criticisms Turing responds to are a theological argument, a heads in the sand stance, a mathematical objection, a consciousness argument, the creative argument and the informal behaviour argument.

The theological argument is to answer the question whether the souls that make humans think according to the christian belief are not part of animals and machines, and therefore machines cannot think. Turing solves this dilemma by arguing that god could give machines souls if he wished to. This argument seems obsolete.

The heads in the sand stance is more interesting. Although it does not involve many arguments, the point is that it would not be good for machines to be able to think, therefore we should believe they don't. Turing does not refute this argument since he does not consider it to be important enough. However an interesting clue in this objection is that the idea that machines can think is somehow disturbing. It might be that this is the basis for objections we dealt with before like that consciousness is something of the organic.

The mathematical objection is the objection that we could ask the machine recursive questions that it will necessarily answer wrongly or incomplete. Turing answers that also humans can answer questions wrongly, therefore the satisfaction from knowing that machines will necessarily do these questions wrong is not very significant. Furthermore, one can doubt the validity of this mathematical argument in current day computers. What has improved significantly in the past 60 years are the capabilities of calculation,

code checking and multithreaded applications. In a sense this problem can be avoided and that more human-like answers can be given even in these kind of situations that are beyond the computational powers of computers. For example running the question in a separate thread while another thread waits for the answer, or if that takes too long gives up, seems quite similar to what a human would do.

The argument from consciousness is another interesting argument, as it is an argument we dealt with before in this thesis. The criticism is that unless a robot can really feel that it is thinking, have pleasure and so forth, that it is not really a thinking machine. Turing's criticism is similar to what was argued for before in this chapter. The only person we are sure that can think and feel is ourself. We could also doubt other humans, but we don't. Therefore we should judge machines by this same standard. The tests purpose is to find a way to test whether machines would be able to think without going into the questions of consciousness. If a machine can get through the test successfully we still are not sure the machine has consciousness, but what we do know is that it can hold a conversation as well as a human can. That is perhaps even more than we would require of a robot in order to have a social relation with it.

The creative argument is that machines necessarily will not be able to make something new while we humans can create new things. Turing initially counters this by saying that we cannot know whether something a human made is really new or just a product from what he learned around him. Still the critique still stands after that remark, since indeed computers can only do what we have programmed them to do. Which in essence means that nothing new is being made after that fact. However, Turing remarks, it is not right to assume that every consequence of the computer has been considered while building it. Which means that even though something might still be a product of human programming, the result is something even that programmer did not anticipate. In that way, machines might still be able to create.

Similar to the creative argument is the informal behaviour argument. The argument that states that unlike machines, humans can adapt and do not need rules in order to cope with situations that are unclear or never accounted for. Machines however, necessarily have a set of rules that governs their behavior. This means that it is impossible to make a robot sufficiently intelligent, as we would be unable to program enough rules. Turing counters this argument by stating that we cannot be sure that we humans indeed do not have a finite set of rules that cope with any situation. This counter is not sufficient because the point is that it is extremely impractical to account for every single event in code, even if there is a finite set. It seems that Turing underestimated this problem, and so did many computer scientists of that time. However, a newer invention of neural networks might help solve this issue, as it seems to be possible for

robots to learn categories of objects in order to help them cope with things they have never encountered before. Even if that path ends up without success, we can still have machines and robots that are not equipped to handle every single situation, as it does not seem to be a requirement that the robot or artificial agent for the purpose of a social relation has to be capable of everything.

There have been a number of cases of computers passing a weak version (Loebner prize) of the Turing test, for example ELIZA. There is however some controversy regarding these passes, as it could very well be the case that the interrogators in question are not skilled enough. This is in part because Turing did not specify the skill required by the interrogators. What is interesting from these passes however is how simple a program can be in order to fool interrogators. Making typing errors for example can play a large part in convincing that one is human, even though programs designed for the Turing test make these errors (on purpose) as well. It is in these details that the difference can be found. A strong Turing test, a test where an interrogator can ask any question could be better.

A strong argument against using the Turing test is that by French (1990). He argues that the Turing test does not test for intelligence, but rather for human intelligence which is true to a certain extent. He argues that there are several types of questions that artificial agents cannot answer correctly, especially those requiring experiencing the world as a human did. French uses an interrogator that can compare the average scorings of certain associations and experiences with a separate group of people from the same culture and use that to pick out the computer every time.

The problem with French's criticism is that he allows the interrogator to have a number of assumptions about the human he is dealing with. He suggests that a computer must try to imitate the culture of the human that is also being interrogated. This is also not required as we would then undermine the Turing test itself, since it might then fail to recognize a case when two humans are put against the interrogator. His methods of determining whether something is human or a computer would not work if we subject the interrogator to a test as well. For example, it must not fail to recognize humans as human too often, for example by putting two humans against him (without him knowing that it is an option) and requiring him to fail to recognize the difference. Preferably this test should be done with two people from vastly different cultures and mannerisms, or intelligence. This way he cannot solely use predetermined average scorings to identify a computer, as two persons also score differently. Perhaps it would even be wrong for the interrogator to presume that all humans in the test are capable of vision or hearing, have had a significant lifespan (for example, use a child as the human counterpart) or are not caring about certain experiences (apatheia). Another option is to have two computers

fight it out from time to time, requiring the interrogator to identify both as computers. If an interrogator can recognize a computer without all these presuppositions about the culture and capabilities of the human being questioned, we can say that the computer failed the test. No honest, capable human being doing his or her best should fail the Turing test.

4.5.2 Virtual Worlds

If it is generally hard to distinguish humans from machines when interacting through a text interface, we might have to take a look at online chatting on sites like Facebook or Omegle or for example in video games. Søraker (2012) argues that social relations (primarily friendships) are not necessarily worse than traditional real world relations. The worse in the previous sentence is referring to subjective well-being of the persons in the virtual relationships. Would they be happier if they had a real friend instead? Perhaps, Søraker argues that it depends per person as the most important difference between virtual and real friendships is that virtual friendships have fewer opportunities (less senses available) to contribute to well-being than real friendships. However, not everyone is equally good at making friends in the real world such that the lowered barriers in the virtual world help such persons making friends. Some (virtual) friends are better than none right?

If we apply that mode of thinking on relationships with robots, we could argue that robot-relations certainly have value. The amount of value is likely even higher than that of virtual friendships as there are more human senses available and therefore more ways to contribute to well-being. However, the importance of self-disclosure (voluntarily or not) means that these robots need some way to stimulate self-disclosure. Therefore it would really help socially if robots would have some kind of system that allows them to recognize facial expressions and emotions to a certain extent. Affectionate computing, a research area dedicated to recognizing and replicating emotions in robots, is busy creating robots with that ability. It is no easy task to recognize emotions, but methods usings facial data and voice recognition (Cowie et al., 2001) have been proposed. Note that even if these systems do not perform as well as one would hope, it is not necessarily a problem, as even humans often fail in recognizing the correct emotional state. If we were to apply a text based interface (as would often be the case in virtual friendships) a way to read through the lines would be needed to detect emotions. Perhaps this is easier than it sounds, as Søraker already mentioned; information like frequency of visits, the time spent online and the time it takes for a person to reply can give cues as to their emotional state. A significant variation from normal behavior could prompt the system to inquire into the life of the person, which might help start some self-disclosure.

Even if it is the case that robots are unable to induce significant self-disclosure from humans, robots can still form a social relationship. Many role playing games have a set of characters played by an AI or more likely just a set script. Often these characters play an important role in the story, but also have their own background story and motivation for helping the player. In that case it is the artificial agent that discloses itself to the player. Even though the player is unable to disclose himself to the character, one still starts to care for it. A similar argument can be made for books. In books it is quite clear that the reader has no influence on the character and that the characters are the only ones that can disclose themselves to the reader. Even though, many people are emotionally touched by characters and their story.

As was argued in Chapter 3, the role of stories in artificial agents and robots is significant. If a story alone can induce an emotional reaction, what are the limits of an interactive robot with its own story? Even though characters in stories and virtual worlds in a sense deceive readers and players of their true nature, there is no reason to blame them for it. A similar argument can be made for robots. Even if robots inherently deceive their users of their intentions, friendship, love and other emotions. As long as they appear to possess these properties it should be fine.

4.5.3 Appearances are enough

Coeckelbergh (2010) argues that genuine emotions, consciousness and states of mind are not necessary for being perceived as having such. Rather we could develop quasi-moral robots that appear to be moral rather than having genuine morality.

Asimov and Reilly (2008) were amongst of the first to describe robot morality. In his Laws of Robotics Asimov argues that all future robots should follow the following three laws.

- First Law: A robot may not injure a human being or, through inaction, allow a human being to come to harm.
- Second Law: A robot must obey any orders given to it by human beings, except where uch orders would conflict with the first law.
- Third Law: A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

The question that comes up with these laws is whether or not following these rules blindly is to be considered moral. This is because even though following the rules might result

in morally acceptable behaviour, it is not done for the right reasons. Morality should, according to the 'emotion view' come from having the right emotions that cause one to act upon it. Having the emotion 'regret' after an action because it leads to pain or hurt of another human is different than evaluating the situation purely on rules and structures and concluding how the least harm is done according to some maximin theory. In such situation there is no regret or even if there is, not for the harm done, but rather for the inability to maximize the outcome. From the point of the emotion view, robots can be considered a sort of mechanical psychopaths. Agents with no consciousness, morality or any sort of emotion that is required for their form of moral behavior. Therefore we should not proceed to make robots like these at all. Who would want mechanical psychopaths?

While it seems that according to the emotion view robots do not have any emotions, states of mind or consciousness, and most likely will not have so in the near future; it does not necessarily result in the conclusion that we should therefore not build robots anymore. We should not think of robots as psychopaths, as that would imply they have a form of consciousness. Robots, in that sense are nothing more than rules and evaluation functions. Building a robot that seems moral is sufficient, as then the rules and evaluations it follows and makes seem morally acceptable to us. Programs, in the end, do not have choices and it is wrong to think of them in that way. Just as a mathematical equation cannot choose the value of x, neither can a program choose the outcome of its execution. A 'right choice', morally speaking, is impossible for a robot.

As Coeckelbergh argues, a robot is not only mindless, meaning they have no consciousness, awareness of themselves or a 'self'. They are also bodiless. The robot-body we see is completely different from how a robot would 'perceive' it. This means that unlike us, they cannot perceive their body as a part of their cognition. In the 'feeling' theory emotions are the way we become aware of our bodily changes. The mental state of emotion follows the bodily feeling, rather than preceding it.

Furthermore, even if a robot would be able to claim that it is conscious, how is it going to prove it? We could let it pass the aforementioned Turing test, but what we would really ask is: Why would we need to prove consciousness? If we take a the path of Descartes, only I know I am consciousness, that only I exist, that only I think. All else is speculation from my part. However I have absolutely no trouble believing my friends, my family and almost all humans (there might be an unconscious human every now and then) are conscious and thinking. There seems no need for us to prove that to each other. As Coeckelbergh argues, why would robots then need to prove so? Isn't the appearance of being conscious enough?

The appearance of consciousness is a rather difficult concept as well however. Something might appear to be conscious to you, but not to me for example. In the simplest example, a Furby toy, might seem conscious to perhaps a child, or an animal. It behaves, it moves, it makes sound and reacts to its environment. It even suggests to have mental states, such as happiness, sadness and being bored. A closer look at its workings suggest a quite different way of interpretation. It is far from conscious, it reacts to certain sensors, it has timers and hard-coded responses to certain sensory information. Its designer would quite likely never see such a robot as conscious. We can alter Floridi and Sander's (Floridi and Sanders, 2004) Levels of Abstraction to fit these levels of interpretation. If we have more understanding of the inner workings of robots, programs or even the mind, then we are less able to see what part of it is conscious. In neuroscience this seems to happen more and more as well. For example, Dennett (1994) argues that humans are just very complex robots, and that in the end, only our complexity allows for our consciousness. Whereas normally we would contribute object recognition to our conscious mind, a neuroscientist might just think of it as the complex operation done in a large area in the back of my brain. If we would know exactly how the brain works, would we lose our conception that we are conscious? If not, we must admit that robots could, in principle, be just as conscious as we are.

Even if we cannot speak of such consciousness in robots, what matters is that they become advanced enough that we might perceive them as having consciousness. When that happens, not only morality, but also friendship relations might come in the realm of possibility.

Coeckelbergh also discusses the apparent inequality present in the way we relate to robots. If robots are indeed, on terms of reasoning, below the average human, we could view them as having the same status as 'disabled' human beings. It does not seem that we have any ground, based on appearances, to deny them this status, on the basis of the argument from marginal cases.

Another option is that of slavery, where the human will always be the masters, and robots, less than human slaves. In history, slaves were also not considered to be human, and could therefore be treated differently. This way of viewing robots might be fruitful. Slaves are also property - Viewing robots just as property, like for example my mobile phone is my property, can allow robots to be treated with some respect, as one does not simply destroy another's property. The problem is that if robots were completely self-reliant and free of any human relation, then there would be no protection from violence against them. These ways of viewing are there so we can treat them with the respect they require as agents with a form of human morality.

4.6 Summarizing

In this chapter was showed that the difference between humans and robots is primarily in the complexity of our being. Even if it is impossible for us to create something as complex as a human being, it is still unclear how much of our complexity is actually important for us to appear conscious and seem to posses a free will. In the next chapter we will discuss whether current and near future robots can have enough complexity to appear as such. Furthermore we will discuss whether these appearances can involve a sufficient amount of affection, usability, interaction and admiration to have a certain social relation with such a robot or artificial agent.

Chapter 5

Our social relations with robots and artificial agents

5.1 Introduction

We have come a long way since the formulation of the research question. The whole thesis is built around the question of what type of human-robot relation could possibly exist in the near future. In Chapter 2 we have seen that a human social relation can be described in the social relation framework described using the properties of Affection, Utility, Interaction and Admiration. In Chapter 3 we discussed the current and near future capabilities of robots and artificial agents. Finally in Chapter 4 we have discussed to what extent these abilities are useful and how we can test these artificial agents and robots for these properties.

In this final chapter we discuss how the capabilities of current and near future robots map onto the social relation framework to finally answer the question whether friendship between humans and robots is a real possibility or if there is some other type of social relation that describes that relation. We will do this by examining each property of the framework separately and see what the limits of current and near future robots and artificial agents are per property.

Let us briefly recap what we have determined in the previous chapters. We determined that the near future robots and artificial agents are robots that do not possess a real consciousness, emotions or a mind as we generally understand these in human terms. However, these robots do possess the capability to converse in natural language, have the capability to move, point and do a certain amount of physical tasks. Furthermore they can keep track of appointments and remind you of certain information at times.

For this robot we can assume that the best capabilities off all existing technologies are present and improved. They can appear to be moral agents, and can appear to have some form of consciousness or emotions, even though they do not posses it truly. The social robot can even have a background story it can use to place conversations in the context of certain aspects in his or her own story, this should make the robot more interesting for the long term. Even a more limited robot like this is incredibly complex to make and we have determined that this is the limit for the near future.

Now we shall explore the robot capabilities in light of the framework, starting with Affection.

5.2 Affection

One of the most important properties for any close social relation is affection. It is also one of the more troublesome properties as it immediately touches upon the problem of appearances versus reality we discussed in the previous chapter. Since symmetry is an important part of social relations we need to discuss both the human-robot affection and the robot-human affection. The problem of appearances versus reality is most prevalent in the latter relation. Before we start discussing the side of the robot however, let us discuss the side of the humans for this property.

What is it that would breed affection for a robot in humans? We have seen that philosophers argue that a wide range of things can cause affection for other humans such as similarity, empathy, attraction, beauty and care. Humans can also have affection for objects, for example for their functionality, beauty or the memories they bring with them. Ihde (1990) and Verbeek (2005) argue that the relation we have with technology like robots is that of an alterity relation. An alterity relation means that a relation with such object is not because it is used to observe some other part of the world but rather as a relation directly with that technology itself. Examples of such relations are the relation one has with a musical instrument or perhaps with one's car. I care for my guitar in more ways than just a musical instrument, I would almost treat it like a person, a quasi-other in Verbeek's terms. In Normans (Norman, 2007) terms this would be caused by my emotional attachment to the instrument caused by the 'shared experiences' and the memories that are connected to the instrument.

We created a separation of two types of affection. One applicable to material, the other applicable to humans. This raises the question whether affection for robots is of the material kind or the human kind. An interesting consequence if it were affection of the material kind is that it becomes irrelevant whether the object or robot has any affection

for the human as well, since we do not expect objects to reciprocate. For the human kind of affection a more symmetric relation seems preferred, although it is not required.

Robots can be subject to either kind of affection, depending on the way the robot is designed. It is easier for the designers to go for a more material affection as this will not require an emotional or conscious appearance, the robot would be purely functional for nonsocial tasks. Nonsocial tasks could include tasks like cleaning the house, provide care or be a clerk at a store. Note that these nonsocial tasks can require conversation with humans in natural language, although always in a strictly business-like manner. However, robots that do need to perform social tasks, such as social-robots, will require a form of humanlike affection. Social tasks are about social interaction for the sake of interaction, for example to combat loneliness. In a sense, the term, humanlike affection is wrong, as not only human-human relation fall under this type of affection, but also human-animal relations. The affection one has for a dog, for example, is of a completely different kind than the affection for any object. The difference is that animals seem to have the capability to reciprocate affection. This ability allows them to fall in the same category of affection as the human affection. Robots can aim for this same affection as animals receive. Paro, the social-robot seal used in a care context for example used this strategy. Several M.I.T. robots also try to achieve this affection by making robots resemble cute (fictional) animals with many ways to express emotions.

The level of affection humans can have for such robots can be high. This is because robots and animals can summon the same type of affection in humans. It seems to be safe to assume that humans can have a high level of affection for animals. This means that robots can also attain this high affection at some point.

Samani et al. (2011) did some empirical research in re-creating feelings of love and affection so that it can be applied to human-robot relations. It should be noted that the love Samani et al. pursue is not the love as we would relate to a spouse but rather that of the love for family and friends. This is close to what I defined to be affection in the framework. Samani et al. designed a robot in order to create a sense of affection in robots. It seems that designing for affection is relatively easy. Small and light things are more likely to be perceived as cute, while the color white induces a sense of trust. The shape of eggs (or infants) apparently plays on our instinct of care and so forth (Samani et al., 2011). It seems that designing a cute robot kitten would strongly affect our affection for such a robot. Although it should be noted that there are some differences between men and women regarding this affection and that the capability for customization of the robot is generally a good idea.

Our inquisition into affection is only halfway done now though. Not only is it important that humans can have affection for robots, the robots also need to have some level

of affection for humans. A robot with real affection seems not possible in the near future since it would require emotions. What is much easier however is a robot that appears to have affection for humans. This can be shown in many ways such as through supporting (through speech or actions), caring and being attentive (by remembering previous encounters and activities). Furthermore such a robot needs to do this for only a limited number of people, so that those people feel 'special' in that regard.

The appearances of such robots would definitely allow for showing a lot of affection. There is however one problem. The framework is based on the feelings of the other person, not how that person appears to the other. We solved that by stating that the appearances are enough to go on, as that is also the only information we can go on. This would be enough if it were not that the actions and appearance of the robot now not only lay in the domain of the capabilities of the robot, but rather in how that robot appears to the user. All the apparent affection in the world might not create the feeling of being appreciated or cared for by someone. Similarly only a small amount of apparent (visual) affection might be enough for others. Therefore the answer to the problem of affection ultimately revolves around the human in the social relation. The capabilities of a robot can only make it more likely that someone finds it affectionate. Capabilities as such make it more probable that there will be a number of people that feel the robot is highly affectionate for them however.

Since robot-human affection is now in the hand of the way it appears to a (perhaps delusional) human, it means that for affection a human-robot social relation can have all the possible values. Therefore the robot-human level of affection is the value 'irrelevant'. To strengthen this point there are examples of humans falling in love and marrying fictional characters or dolls. This 'object sexuality' as it is called seems to be a relation where the affection from object to human is completely irrelevant. One could bite the bullet and call object sexuality misguided. That object sexuality is not genuine real love, but rather a form of extreme obsession. One might be right about that, but in the case that the feelings of the human in that relation are like real love, then it would be a relation where the object-human side of affection is irrelevant.

5.3 Utility

The second property is that of utility. Many social relations are based on this property. Social relations like the business relation, buddies and companions are rooted in this property. An exception is friendship, since that social relation is notable for its irrelevant utility. Friendship is beyond the need for use, even if it is quite useful in many occasions.

We have always had a peculiar relation towards objects, instruments and artefacts in our world. Apart from the alterity relation Verbeek and Ihde (Verbeek, 2005) would argue that all relations with technology are a way to interface with another part of the world. Norman (2007) would suggest that one of the most important aspects of any technology is its functionality.

Technology is always created with a certain goal in mind. Even art can be considered to have a functionality in the form of invoking emotions or inspiring idea's in people. Even the most seemingly useless object most likely serves or served some goal, being physical or emotional in nature. So too must a robot have a goal for the persons investing in it. This means that a robot will always have a purpose, however small, for its owner. This means that its utility will never be completely irrelevant, since it is for this reason that the investment is made. Note that this does not immediately rules out friendship, since the required utility in one type of relation does not exclude the possibility of irrelevant utility in others. Let us first look at the utility for a person.

How much utility can a social-robot have for a person? Well that mostly depends on the importance of the task performed by this social-robot. We should note however that it is quite likely that any task performed by social-robots, even if quite versatile, can probably be solved by technological means in simpler ways. The notable exception to that is social relations in certain circumstances. Therefore the type of utility up for discussing should be social in nature. Utility in the form of pleasurable experiences, utility as a way to combat loneliness. The utility of other types of robots is not so relevant, as many artificial agents and robots are designed with a clear goal and purpose, a good performance in those area's will provide them with a high degree of utility almost naturally.

The level of utility of social robots however is harder to pinpoint. A very lonely person is probably quite helped by a social robot that relieves him or her of that. Robots used to combat social-anxiousness or other social disorders are perhaps even more useful in the long run.

The highest level of social relation, a true friendship, needs a utility relation of a whole different kind. There would need to be no apparent goal for the relation, since the utility needs to be irrelevant. If there is no reason for a social relation to a robot, why would you start such a relation? This brings us to the similar question we faced in Chapter 2. Why do we need friends, and how do we make them? Apart from the shared activities one would need to grow into a friendship from a certain point in a more formal relation. From the utility perspective a more formal social relation based utility is certainly a possibility. The problem arises from the moment that one would want to expand the current social relation to a friendship. At that time the real purpose of the robot would

need to disappear and while the human in the social relation would still want to relate to the robot. The problem then is that the robot fulfills a purpose it is no longer required to do, and might even have to be able to adapt and change the way it relates to that person.

Another problem with a friendship relation is that the irrelevant utility is both ways. Which means that even if the utility from human to robot can somehow become irrelevant, also the utility from robot to human needs to become irrelevant. This implies that the robot must be fully self-sufficient and independent in all its resources with regard to the other person in this social relation. This means that either the robot is free and has its own life somehow, or is owned by somebody else. Quite a problematic requirement in the current consumerist society. From the utility perspective it is therefore unlikely that one can be a friend with a robot he or she owns.

It seems that even though the social relation between human and robot on the criteria of utility can have any value, it is impossible for any side to have no relevance from this property at all. From the point of view of the robot it could be possible, but only if it is a free robot or if the robot is owned and maintained by a third party that does not influence the relation with another human in any way. Both options seem quite unlikely now and even in the near future. The first because a 'free robot' would be quite pointless to a certain extent as it would not have anything to do in its 'free time', furthermore it seems strange for anyone to design and produce robots that have a notion of being free, as the company that produces said robot would not be able to sell the robot. The second is unlikely because it would require a third party to maintain a robot that henceforth has social relations with persons completely out of its intended circle. Furthermore the company is not allowed to ask for money or have any form stake in those relations as that would make utility relevant again.

5.4 Interaction

The interaction property seems to be an easier property to assign to a robot-human social relation. Quite so because social robots can be designed to perform well in interaction. One of the biggest challenges is to create robots that can converse in natural language and preferably stay interesting for longer periods of time. It is from the wish to create robots that can talk like us, that the notion of robot friendship seems to have emerged. While it seems that social robots used in research seem to mainly focus on this property, it is far from clear whether the capabilities in this area will be sufficient for good interaction.

The capability to converse in natural language is quite important and in the previous chapter we discussed the difficulties and problems associated with this capability. Also the ability to use bodily language and facial expressions to convey a message is being worked on, although a convincing and correct use of these expressions is not easy. Even if all these barriers are overcome and an artificial agent or robot can actually converse acceptably well and talk for longer periods of time at a high level there is still more to interaction than just that.

Interaction includes more than just talking and communicating. Also shared activities and mutual interests play a large role. Lewis (1960) argued that friendships form from mutual interests and shared activities. One way to find people to converse with could be to pursue that interest and meet people with similar interests. Whether this chase includes an education, club, travel or video gaming, there will be many ways in which one can find persons with similar interests. How does this work for robots though? Even if we stick to just conversing, robots would need some way to be interested in things and not interested in others. If all the interest comes from the human one could doubt the social relation would last long. It would really help if robots would be able to show enthusiasm for some subject. French (1990) would argue that artificial agents and robots will never be able to do this effectively since robots would need to experience the world the same way we do to effectively make interesting connections. This would mean that even if robots could converse acceptably, the content of these conversations would be quite boring and predictable. Adding a story as discussed in chapter 3 might help somewhat relieve this burden, but might not be enough for social relations for long periods of time.

So the extent of which we can converse with robots is probably quite limited. Also our ability to adapt to any sort of activity might overwhelm a robots capabilities, for example learning and understanding a new board or card game. In such situation a robot would not only need to derive the rules of this new game from spoken language (an incredible feat that might perhaps be achieved in the near future) it must also produce a model of the game and learn how to effectively construct a strategy for this new game. Perhaps this is not as hard as it sounds but the learning and adaptive capabilities of the social robot would need to be incredible if it were to keep up with humans. Formal conversations, especially those in a certain context, are much easier and often only require only short and none-creative answers. A medium level of interaction is certainly within reach.

A high level of interaction seems just out of reach. Perhaps that a higher amount of lower quality interaction could substitute for a human to human conversation, but any type of long term communication with a robot stays limited. One could argue that there

are humans that are not capable of having certain interests or are in conversational capabilities equal or less than these artificial agents and robots of the near future. This might indeed be the case, and there are two solutions for this apparent problem. One could solve it by suggesting that indeed robots and artificial agents can have a high level of interaction, because, they would argue, all speech capable humans will have a high level of interaction with friends and family. However, one could just as well argue that humans with lesser capabilities are unable to have interaction on a high level which means that their relation with friends and family is different than the way defined in chapter 2.

Contrary to the other properties, this one is always symmetric, so if the human to robot is at maximum a medium level of communication, the other way around is at most medium as well.

5.5 Admiration

Admiration is perhaps the most troublesome property for human-robot social relations of all properties. This is because admiration has relevance to virtues, morality and equality. All of which properties that are not easy to apply to robots and artificial agents. As we discussed in the previous chapter, appearances are enough for morality but also for virtues and equality. However appearing equal and virtuous are not quite easy. Let us start with a brief exploration of the virtuous appearing robot. So when would a robot appear virtuous? Well if we go by the virtue ethics of Aristotle (Aristóteles, 1991) then the robot must act according to the golden mean between two vices, one vice of excess and the other of deficiency. An example of such a virtue would be courage, and its vices would be cowardice and rashness. The problem of virtues and robots is twofold, Firstly how would a robot display such virtues? Secondly how would a robot recognize virtues in another?

A robot and artificial agent can have problems displaying virtues because these virtues will perhaps be incomparable with our human virtues. For example what would a courageous robot need to do? Sacrifice himself to save a human life? If it did wouldn't it be considered normal for robots to do so? If robots are all programmed to be courageous in that sense how would such acts be perceived by humans as such? The robots of the near future will not be able to escape the prejudice (and rightly so) of being a machine. Which means that they will be less capable and less valuable than a human life. If that is the case, a robot sacrificing itself to safe a human life would not earn him much credit. The manufacturer would get most of the credit.

The other way around, the ability for robots to appreciate the virtues of humans seems way beyond anything a near future robot would be able to comprehend. Most likely such robots will not be able to differentiate between good and bad, perhaps at most between persons that harm the robot and persons that take care of the robot. If there would be a (for humans) clearly bad person, but that takes good care of his or her robot it would seem to that robot that the person is a good person, even though he treats other people terribly for no reason. It is more likely that all near future social robots will not have any clue of morally right or wrong actions done by human agents. At most it will not execute certain commands that go against some in-built ethical principles that protects the people around it from harmful actions the robot could take.

Another problematic point for human-robot relations regarding the property of admiration is the fact that robots are generally owned. A relation with an owner is an interesting one, as it can take many forms or shapes. A dog is owned for example, and it would see its boss as part of its litter or kennel. This relation is rather positive in nature generally as the dogs are happy to be a part. For human-human social relations regarding ownership (generally known as slavery) is met with a much more negative connotation. Although for the slavers at the times it was not really a human-human relation, but rather the slave was seen as something less than human. Apparently not worthy of the same respect, dignity and freedom a human deserves. Human-robot relations have a bit of both. In a sense robots can be more like a slave than a dog in the sense that is can talk and communicate better, but in another sense a robot is more like a dog than a slave in the way it is aware of its owner and the way it will view that person. The problem is thus that because of the ownership a robot is closer to a slave or dog in the admiration it receives. Is there a way to circumvent this problem?

We could explore the possibilities of free robots, but as was argued in Utility the freedom of robots does not seem like a fruitful and likely possibility. Another option would be to change the way we view robots. Perhaps we can treat robots like we would treat a child instead of a piece of property this might be possible. However empirical research shows that humans often have trouble accepting children of other parents as one of their own (especially if another child is present) so the chance that a robot can fully enjoy that privilege is quite slim (Mekos et al., 1996).

In a sense this means that robots will never receive much admiration. Even a medium amount of admiration seems out of bounds for a robot as that is the amount of admiration one ascribes to a person of equal standing. A human to robot admiration level for social relations will always be low. The other way around is different. The admiration a robot shows to his owner can be at various levels, but robots will not be able to properly

identify virtuous acts or virtuous persons. It will only care for its owner and for it the amount of admiration is quite irrelevant.

This means that admiration in social relations between humans and robots is quite unique in that it does not seem to comply with any current human-human relation. The only human-human relation with a low admiration one way and irrelevant admiration the other way would be the master-slave relationship.

5.6 Conclusion

In this chapter we saw how the combined results of the capabilities of robots, the way social relations are constructed and the philosophical background of moral appearances resulted in a maximum strength of each property in the social relation framework. One of the most important conclusions we can draw from this is that robots now and in the near future cannot be real friends. We saw this was because the properties utility, interaction and admiration all cannot get to the level we determined is required for friendship.

From the utility point of view the problem was that because robots are owned and cannot have a consciousness or free will they can never have an irrelevant utility towards someone else. Robots are always constructed and brought into someones environment with a purpose. This purpose could very well be to keep the person company and prevent loneliness in that sense. A robot could indeed be used as a surrogate for a friend, unfortunately that does not give a robot the ability to be an actual friend on itself.

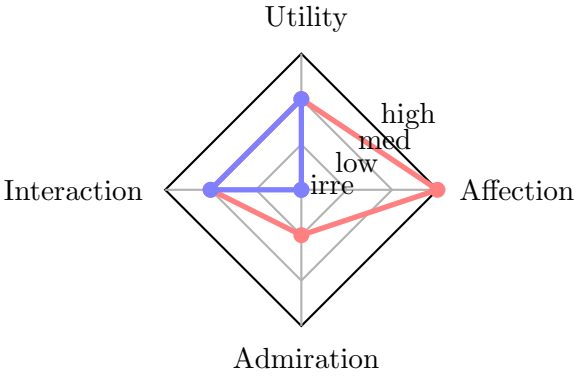


FIGURE 5.1: Human-robot in red and robot-human in blue

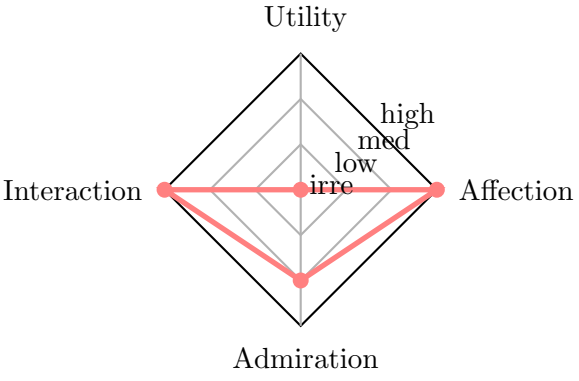


FIGURE 5.2: Friendship relation.

From the interaction perspective we saw that robots cannot be friends because they could never match the experiences and interests required to not only communicate fluently, but to actually share an experience, activity or engage into a mutual interest. This means that at best a robot could try to be entertaining and interesting for a shorter period of time. A self and experiences seem required for a more interesting and ongoing social relation.

Finally, from the admiration point of view we saw that the fact that a robot is constructed, owned and programmed causes us to be unable to perceive virtue and equality in robots. Instead we see a machine made for certain tasks and our admiration, if that task is completed satisfactory, will be directed to the producers and programmers of the robot.

The only thing going for robot friendship is affection. It seems that robots, similarly to pets, can be granted a high amount of affection. Affection alone is not enough for friendship however. There are several other social relations that have a high amount of affection. Perhaps we can try to discover what kind of relation with a robot does seem possible.

The admiration property seems to suggest a sort of master-slave social relation. We could categorise a master and a slave as having irrelevant affection for each other. This is similar to that of a human and robot, as a human can have any amount of affection for a robot it does not particularly matter for the relation. Also at the utility property the master-slave comparison holds as robots can be constructed at any utility but irrelevant, a slave in that sense has a similar range of utility. A slave can be anything from quite useless to highly efficient, but it always matters to the master. Reversely, from robot to owner also holds. Since both the slave and the robot are reliant on their master or owner to provide them with the necessary resources for survival. Finally the interaction property also suggests that a master-slave kind of relation seems the best way to describe a human-robot social relation. Although there are many types of social relations that require only medium interaction, a master and slave relation is one of them.

So, it seems that a master-slave relation is the best way to describe the relation between human and social robot. Note that this is not just merely any robot, but specifically a robot designed to be as interactive, playful and affectionate as possible. Even a relation with that robot seems to result in a master-slave relation.

This all sounds quite negative, since we have a negative connotation with slavery, especially the racism and abuse that is related to the practice. Furthermore it seems wrong now on many levels to take the freedom of any human. This thesis even added another reason, because apparently you cannot be friends with a non-slave as a slave, which

would significantly reduce the number of people one can be friends with. Even though we named the social robot - human social relation a master-slave relation it does not mean it entails all of the negative connotations that slavery brings, for the very reason that this time, they really aren't human. Even if the robots are abused and we discriminate against them in ways it does not matter as much. We do the same to many objects and technologies. How many cars are vandalised every day? These are also definitely objects people can care about. That a robot exhibits a form of apparent intelligence and consciousness does not mean it actually possesses it. Perhaps the destruction of a robot is a far greater crime than the destruction of a car in the future, but that does not mean that it will ever be held equal to taking a human life. Rightly so.

5.6.1 The value of Robot-Human relations

As we have seen, genuine friendship with robots is not on the horizon. However, that does not mean that there is no future for the social robot. Even if a relation with a social robot can never be as strong as the relation with a friend, this does not mean there is no value in these different relations. Especially when too little traditional social interaction and friendship is available for a person it can definitely be a relation that has a lot of value. A relation with a social robot does not have to compete with friendship to be useful, as we could still use the social robots in cases where friendship is not available. The subjective well-being of the persons in question can be improved using social robots.

What this thesis does show however is that robots are not a good replacement for friendship. It could go a long way but if a genuine friendship is an option, then that is certainly preferred. The goal of making a robot friend is productive, even if it is not achievable. Technological advancements that can enhance the quality of the social relation can directly impact the quality of the 'simulation of friendship' that robots perhaps can provide. It has been shown that a simulated trip to a waterfall is not as beneficial as a real trip, but it is definitely better than no trip at all (Søraker, 2012). If that works for something soothing as visiting a waterfall, why would that not work for the good effects of social relations? It seems that social robots can still play an important role in the battle against loneliness in the future.

One important point of discussion we have neglected so far is the role of outside observers on the relation between humans and robots. We could imagine a world where some men chose a 'female' robot companion for their sexual needs. From the perspective of the person it might seem as a suitable solution to a problem they face. Even now sex dolls are a business one should not underestimate, so having social robots with similar

features does seem like a distinct possibility, even if many of us might find that repulsive or morally wrong.

What seems normal for one person might seem awkward or morally wrong for a large part of the community. Turkle (2012) used an example of mothers that now pay more attention to their cell phones than their children. It seems like bad parenting to us, but what if this role will be fulfilled by robots in the future? What will be the public opinion on the use of social robots in many places we would normally find a human agent? What if all stores and McDonald's staff are replaced by robots, social and not? From the capabilities we saw that robots will be especially adept at short businesslike communication, especially if contained within a certain context, so robots like this are certainly a possibility.

It is important to understand that the existence of advanced social robots can have a way greater influence on society than just one on one social relations. Our concepts of conversation, relating and the way we use technology might change. We might end up finding it completely normal for every family to own a social robot that takes care of the kids and the house, even if that means that the next generation is largely brought up by robots. It might even be that at that time our conception of friendship has changed in such a way that friendship with robots does not seem so far fetched. This thesis already uses a quite strong conception of friendship inspired by ancient literature. The way 'friend' is often used on the internet (for example facebook) does not comply with the framework at all. Many online 'friends' are do not enjoy the benefits of affection, utility, interaction or admiration.

The ways in which technology can influence society and perpetuate into affairs that used to belong to humans only are numerous. Technology has changed the way we relate to friends and has in part also changed the concept of friendship itself. Technology has changed the way we make friends and the way we spend time with them. Who know what changes a widespread introduction of advanced social robots will cause. It does not seem out of the question that we will mostly communicate with and through artificial agents and social robots in the near future.

5.6.2 A glimpse of the future

We have established that real friendship with robots is not possible in the way we understand friendship with other humans. In this thesis we have shown that the most social type of relation is like a master-slave relation. This is in part by the inability for robots to own themselves.

The world Samani et al. (2011) envisioned with robot friends, family members and even lovers is not going to happen in the near future. Naturally there will be people that are highly affectionate for their robot and it seems likely that there will be robots used for sexual activities, but the relation we will have with such robots will not really become that of a genuine lover or friend. What we could wonder about however, is what is then the relationship we will have with these robots? A master-slave relationship sounds quite gruesome but it is only slightly different than a master-servant relationship for example. The difference is that a servant could quit and is therefore treated with higher respect. Task and relation wise however these two map quite well. So what robots will be able to replace well are the servants.

Instead of imagining a future with robot friends, family and lovers, we could imagine a future with robot maids, butlers, shop clerks, phone operators, garbage collectors, cops, drivers and medical staff. Robots that bring you a cup of tea when you want it. Robots that do dangerous and dirty jobs that are relatively simple and repetitive. Surely robots will have to be able to converse to some degree, and perhaps play games and help with other types of entertainment. Robots might be able to teach humans (if the students are willing). Robots could build houses and flats on their own using a blueprint. There will be many amazing things robots can do, but their strengths will be in repetition and simple well defined tasks. The research into lovotics and robotics in general will be helpful for all the above tasks. Creating a sense of affection for robots can also be useful when required for a role. A robot house servant should perhaps induce enough affection to be liked and taken care of when needed. Perhaps that will even make its mistakes fun and easy forgiven.

Naturally there will be robot lovers and love robots. It is quite probable that love robots will be used for sex and companionship, since love dolls are quite an industry already. We should however not mistake these love robots for real affectionate relationships. Robots could perhaps (to some extent) replace prostitutes, but robots will not replace lovers, because robots cannot be independent individuals, share experiences and interests. Some people will not mind, as there are people that are fully content with dolls or even bodiless virtual characters. However this does not mean that a real love relationship does not require a body or the sharing of experiences and interests. We should not confuse affection and eros with a complete social relation. One needs to look at all the properties in a social relation in order to assess the type of social relation. Just seeing a high amount of affection does not mean interaction, utility and admiration are at the right levels as well. Samani et al. seems to have neglected these aspects when they envisioned the future with robots love-relations. They are trying to make robots with the right level of affection, but the real challenge is making robots with the right level of utility, interaction and admiration as well.

5.6.3 Recommendations for design

The research question seems to be answered to a sufficient degree now. Hopefully the thesis has inspired you in a number ways and gave you new questions and thoughts. There are still many subjects we could further implore upon however. One of those is the question of how one would build such a futuristic social robot. Let us briefly discuss some aspects of robot design that could be of use. Most of these are no surprise after reading the rest of the thesis, but it can be considered a short summary and practical application of points made in previous chapters in this thesis.

The appearance of a social robot is quite important for our expectations and our willingness to play or interact. If the robot is all gray and square it seems less likely that we approach such robot with an intention to socially interact. It is therefore key to make a robot appear sociable and cheerful. Just making the robot look like a human is not the solution however. The uncanny valley hypothesis states that something that not quite looks like a healthy human is more likely to be revulsing than sociable and cheerful. Since mimicking human behaviour and capabilities is still well beyond the capabilities of social robots it might be a good consideration to keep clear of a full human appearance. The MIT researchers such as Breazeal seem to use this strategy with robots like Leonardo and Kismet.

A further trick that seems useful to increase social interaction with social robots is giving the robot a form of a personal story. If it appears that the robot has a history and also things going on it gives the impression that such robot has an identity. If the robot has preferences, a set of idea's and topics it can passionately talk about (perhaps itself), people are way more likely to stay interested for longer periods of time. It would be great if there would be some form of response to questions like; "How do I look?" and it could answer with "You look great in red clothes" or "I think blue is better on you". Colour is relatively easy to detect and mostly an individual preference. Even if the advice does not actually make sense it is still a form of identity that is being portrayed. Selfhout et al. (2009) argue that musical preferences play a role in the formation of friendship in adolescents. For a robot it would therefore be key to show a preference (and perhaps adapting to the preference of the human) and to some extent have a similar preference as the human.

What designers should also keep in mind is that humans cannot really have a real friendship relationship with robots and therefore such a relation should not be forced by robot design. It is absolutely fine for robots to be friendly and engaging, but interaction should be initiated and terminated easily by the human. It should not be a problem to just walk out on a robot for example, something we would not do with actual friends.

This way the robot is less likely to annoy someone. Note that this does not hold for all robots. Care robots that are designed to get people to take their medicine should not be shy from initiating conversations and reminders themselves. For care robots it is not their goal to be friendly and sociable. It is more important for such robot to fulfill its function.

Function is naturally important for the design of a robot. It does not seem to warrant much thought. What we could keep in mind however is to what extent social interaction will be required for the functioning of the robot. A robot clerk does not need to be sociable, while a robot designed to combat loneliness is. The specific context in which a robot will operate will influence its design and our relation to it. In this thesis has been shown however that the hard problems of robot social relations are not making human affection for robots, but rather make people have genuine admiration for and the right utility relation with robots.

Bibliography

Aristóteles (1991). *Aristotle: Nicomachean Ethics: Books VIII and IX*, volume 8. Hackett Publishing.

Arkin, R. C. (2008). Governing lethal behavior: Embedding ethics in a hybrid deliberative/reactive robot architecture part i: Motivation and philosophy. In *Human-Robot Interaction (HRI), 2008 3rd ACM/IEEE International Conference on*, pages 121–128. IEEE.

Asimov, I. and Reilly, T. (2008). *I, robot*. Bantam Books.

Aydin, C. (2013). The artifactual mind: overcoming the inside–outsidedualism in the extended mind thesis and recognizing the technological dimension of cognition. *Phenomenology and the Cognitive Sciences*, pages 1–22.

Bacon, F. (1871). Of friendship. *The Essays*.

Bates, J. (1994). The role of emotion in believable agents. *Communications of the ACM*, 37(7):122–125.

Botelho, L. M. and Coelho, H. (1997). Emotion-based attention shift in autonomous agents. In *Intelligent Agents III Agent Theories, Architectures, and Languages*, pages 277–291. Springer.

Breazeal, C. (1998). Early experiments using motivations to regulate human-robot interaction. In *AAAI Fall Symposium on Emotional and Intelligent: The tangled knot of cognition, Technical Report FS-98-03*, pages 31–36.

Brey, P. (2010). Values in technology and disclosive computer ethics. *The Cambridge handbook of information and computer ethics*, pages 41–58.

Cañamero, D. (1997). Modeling motivations and emotions as a basis for intelligent behavior. In *Proceedings of the first international conference on Autonomous agents*, pages 148–155. ACM.

Clark, A. and Chalmers, D. (1998). The extended mind. *analysis*, 58(1):7–19.

Cocking, D. and Kennett, J. (1998). Friendship and the self. *Ethics*, 108(3):502–527.

Coeckelbergh, M. (2010). Moral appearances: Emotions, robots, and human morality. *Ethics and Information Technology*, 12(3):235–241.

Cole, A. (2004). What hegel's master/slave dialectic really means. *Journal of Medieval and Early Modern Studies*, 34(3):577–610.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. G. (2001). Emotion recognition in human-computer interaction. *Signal Processing Magazine, IEEE*, 18(1):32–80.

Dahl, G. E., Yu, D., Deng, L., and Acero, A. (2012). Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *Audio, Speech, and Language Processing, IEEE Transactions on*, 20(1):30–42.

Dennett, D. C. (1994). Consciousness in human and robot minds.

El-Nasr, M. S., Ioerger, T. R., and Yen, J. (1998). Learning and emotional intelligence in agents. In *Proceedings of AAAI Fall Symposium*, pages 1017–1025.

Feldman, F. (2002). The good life: A defense of attitudinal hedonism. *Philosophy and Phenomenological Research*, 65(3):604–628.

Fiske, A. P. (1992). The four elementary forms of sociality: framework for a unified theory of social relations. *Psychological review*, 99(4):689.

Floridi, L. (2008). The method of levels of abstraction. *Minds and Machines*, 18(3):303–329.

Floridi, L. and Sanders, J. W. (2004). On the morality of artificial agents. *Minds and machines*, 14(3):349–379.

French, R. M. (1990). Subcognition and the limits of the turingtest. *Mind*, 99(393):53–65.

Friedman, B. (1997). *Human values and the design of computer technology*, volume 72. Cambridge University Press.

Gadanho, S. P. (1999). Reinforcement learning in autonomous robots: an empirical investigation of the role of emotions.

Gazzaniga, M. S. (1997). *Cognition, computation, and consciousness.* Oxford University Press.

Gockley, R., Bruce, A., Forlizzi, J., Michalowski, M., Mundell, A., Rosenthal, S., Sellner, B., Simmons, R., Snipes, K., Schultz, A. C., et al. (2005). Designing robots for long-term social interaction. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2199–2204.

Goetz, J., Kiesler, S., and Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. The 12th IEEE International Workshop on*, pages 55–60. IEEE.

GonzáLez, Y. S., Moreno, D. S., and Schneider, B. H. (2004). Friendship expectations of early adolescents in cuba and canada. *Journal of Cross-Cultural Psychology*, 35(4):436–445.

Hawking, S. (2013). Does god play dice. *http://www.hawking.org.uk/does-god-play-dice.html*.

Ihde, D. (1990). *Technology and the lifeworld: From garden to earth*. Number 560. Indiana University Press.

Kandel, D. B. (1978). Similarity in real-life adolescent friendship pairs. *Journal of personality and social psychology*, 36(3):306.

Kant (1930). *Lecture on Friendship*, volume 8. Translated by Infield. Hackett Publishing.

Klein, J., Moon, Y., and Picard, R. W. (2002). This computer responds to user frustration:: Theory, design, and results. *Interacting with computers*, 14(2):119–140.

Leite, I., Martinho, C., Pereira, A., and Paiva, A. (2008). icat: an affective game buddy based on anticipatory mechanisms. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 3*, pages 1229–1232. International Foundation for Autonomous Agents and Multiagent Systems.

Lewis, C. S. (1960). *The four loves*. Collins.

Lucivero, F., Swierstra, T., and Boenink, M. (2011). Assessing expectations: towards a toolbox for an ethics of emerging technologies. *NanoEthics*, 5(2):129–141.

MacDorman, K. F. and Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, 7(3):297–337.

Marchant, G. E., Allenby, B., Arkin, R., Barrett, E. T., Borenstein, J., Gaudet, L. M., Kittrie, O., Lin, P., Lucas, G. R., OMeara, R., et al. (2011). International governance of autonomous military robots. *Columbia University Review of Science & Technology Law*, 13.

Martínez-Gómez, J., García-Varea, I., and Caputo, B. (2012). Overview of the imageclef 2012 robot vision task. In *CLEF (Online Working Notes/Labs/Workshop)*.

Mekos, D., Hetherington, E. M., and Reiss, D. (1996). Sibling differences in problem behavior and parental treatment in nondivorced and remarried families. *Child Development*, 67(5):2148–2165.

Modi, P. J., Veloso, M., Smith, S. F., and Oh, J. (2005). Cmradar: A personal assistant agent for calendar management. In *Agent-Oriented Information Systems II*, pages 169–181. Springer.

Montaigne (1958). *Of Friendship, Essay 28*, volume 8. Translated by Frame. Hackett Publishing.

Morignot, P. and Hayes-Roth, B. (1994). Why does an agent act? In *IN MT COX & M. FREED (EDS.), PROCEEDINGS OF THE AAAI SPRING SYMPOSIUM ON REPRESENTING MENTAL STATES MECHANISMS. MENLO PARK, AAAI.* Citeseer.

Nagel, T. (1970). Death. *Nous*, 4(1):73–80.

Norman, D. A. (2007). *Emotional design: Why we love (or hate) everyday things*. Basic books.

Nozick, R. (2012). The experience machine. *Ethical Theory: An Anthology*, 14:264.

Pakaluk, M. (1991). *Other selves: Philosophers on friendship*. Hackett Publishing.

Pennebaker, J. W., Zech, E., and Rimé, B. (2001). Disclosing and sharing emotion: Psychological, social, and health consequences. *Handbook of bereavement research: Consequences, coping, and care*, pages 517–543.

Pizzocaro, D., Parizas, C., Preece, A., Braines, D., Mott, D., and Bakdash, J. Z. (2013). Ce-sam: a conversational interface for isr mission support. In *SPIE Defense, Security, and Sensing*, pages 87580I–87580I. International Society for Optics and Photonics.

Samani, H. A. and Cheok, A. D. (2010). Probability of love between robots and humans. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 5288–5293. IEEE.

Samani, H. A., Cheok, A. D., Tharakan, M. J., Koh, J., and Fernando, N. (2011). A design process for lovotics. In *Human-Robot Personal Relationships*, pages 118–125. Springer.

Searle, J. R. (1982). The chinese room revisited. *Behavioral and Brain Sciences*, 5(02):345–348.

Selfhout, M. H., Branje, S. J., ter Bogt, T. F., and Meeus, W. H. (2009). The role of music preferences in early adolescents friendship formation and stability. *Journal of Adolescence*, 32(1):95–107.

Singer, P. W. (2009). Military robots and the laws of war. *The New Atlantis*, 27:27–47.

Sloman, A., Beaudouin, L., and Wright, I. (1994). Computational modelling of motive-management processes.

Smeets, D., Hermans, J., Vandermeulen, D., and Suetens, P. (2012). Isometric deformation invariant 3d shape recognition. *Pattern Recognition*, 45(7):2817–2831.

Søraker, J. H. (2012). How shall i compare thee? comparing the prudential value of actual virtual friendship. *Ethics and information technology*, 14(3):209–219.

Sparrow, R. and Sparrow, L. (2006). In the hands of machines? the future of aged care. *Minds and Machines*, 16(2):141–161.

Telfer, E. (1970). Friendship. In *Proceedings of the Aristotelian Society*, volume 71, pages 223–241. JSTOR.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236):433–460.

Turkle, S. (2012). *Alone together: Why we expect more from technology and less from each other*. Basic Books.

Ventura, R., Custódio, L., and Pinto-Ferreira, C. (1998). Emotions-the missing link? In *Emotional and Intelligent: the Tangled Knot of Cognition. 1998 AAAI Fall Symposium*, pages 170–175.

Verbeek, P.-P. (2005). *What things do: Philosophical reflections on technology, agency, and design*. Penn State Press.

Weber, M. (2001). Determinism, realism, and probability in evolutionary theory. *Philosophy of Science*, pages S213–S224.

Wright, I. (1996). Reinforcement learning and animat emotions. In *From Animals to Animats IV, Proceedings of the Fourth International Conference on the Simulation of Adaptive Behavior*, pages 272–281.

Wynsberghe, A. (2012). *Designing robots with care: creating an ethical framework for the future design and implementation of care robots*. Universiteit Twente.

Zech, E. (2000). The effects of the communication of emotional experiences. *Unpublished doctoral dissertation. Louvain-la-Neuve, Belgium: University of Louvain.*