Final thesis

Social Dimensions of Robotic versus Virtual Embodiment, Presence and Influence

by

Sam Thellman

LIU-IDA/KOGVET-A-16/005-SE

2016-08-19
Final thesis

Social Dimensions of Robotic versus Virtual Embodiment, Presence and Influence

by

Sam Thellman

LIU-IDA/KOGVET-A–16/005–SE

2016-08-19

Supervisor: Tom Ziemke
Examiner: Arne Jönsson
Copyright

The publishers will keep this document online on the Internet – or its possible replacement – from the date of publication barring exceptional circumstances.

The online availability of the document implies permanent permission for anyone to read, to download, or to print out single copies for his/her own use and to use it unchanged for non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional upon the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility.

According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement.

For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its www home page: http://www.ep.liu.se/.

© Sam Thellman
Abstract

Robots and virtual agents grow rapidly in behavioural sophistication and complexity. They become better learners and teachers, cooperators and communicators, workers and companions. These artefacts – whose behaviours are not always readily understood by human intuition nor comprehensibly explained in terms of mechanism – will have to interact socially. Moving beyond artificial rational systems to artificial social systems means having to engage with fundamental questions about agenthood, sociality, intelligence, and the relationship between mind and body. It also means having to revise our theories about these things in the course of continuously assessing the social sufficiency of existing artificial social agents.

The present thesis presents an empirical study investigating the social influence of physical versus virtual embodiment on people’s decisions in the context of a bargaining task. The results indicate that agent embodiment did not affect the social influence of the agent or the extent to which it was perceived as a social actor. However, participants’ perception of the agent as a social actor did influence their decisions. This suggests that experimental results from studies comparing different robot embodiments should not be over-generalised beyond the particular task domain in which the studied interactions took place.
Acknowledgements

Tom Ziemke, I have felt like a well-watered plant under your supervision. I believe myself to have grown significantly during this period thanks to your perceptive support and warm encouragement. Thank you for teaching me many things.

Caroline Norén, Camilla Kirkegaard, Karl Nygren, Mattias Kristiansson, Annika Silvervarg, Agneta Gulz, and Nils Dahlbäck, you have all influenced my work. Thank you for sharing your thoughts and expertise.

I spent a lot of time together with the members of the Cognition and Interaction Group during the making of this thesis. Thanks to all of you for making the group a great place to be for an aspiring researcher such as myself.

My beloved Sofia Lindvall, you support me always. Thank you for being there for me.
Chapter 1

INTRODUCTION

1.1 Why artificial systems need to perform socially

In 1998, Christiano Castelfranchi wrote:

If we want a computer to be not ‘just a glorified pencil’ (Popper, BBC interview), not a simple tool but a collaborator, an assistant, we need to model social intelligence in the computer. If we want to embed intelligent functions in both the virtual and physical environment (ubiquitous computing) in order to support human action, these distributed intelligent entities must be social to understand and help the users, and to coordinate, compete and collaborate with each other. (Castelfranchi, 1998, p. 158)

Castelfranchi describes a future technology that collaborates with people in social contexts – an artificial system that acts on basis of social awareness. What can we expect from technology that accomplishes this with proficiency? Will it be able to do some of the things that computers already do for us (such as controlling machinery in industrial production lines, organising, monitoring and manipulating cash flow in business and banking, and aiding the design process in various fields of engineering) more efficiently or with better results? Will it be able to engage us, entertain us, enlighten us or take care of us in novel ways?

Agent-based computing research, such as research on interaction between humans and robots or virtual agents, shows that complex artificial systems that perform socially affect people in psychologically and behaviourally

1
unique ways. These systems procure in us a sense of presence – we experience them as social objects that manifest humanness (Lee, 2004). This research demonstrates that people tend to perceive of such systems as agents – fellow sentient cognisers, endowed with intentions, beliefs and desires – rather than as insentient machines. As such, they are judged trustworthy or unreliable, are held socially accountable for their actions, exert social pressure, create social norms and provoke social behaviour in people.

Artificial and natural agents (e.g. humans and animals) alike exhibit behaviour that intuitively appears to us to be directed by goals and guided by reason, that is intentional behaviour. We observe such behaviour both in entities that we are convinced have goals and are capable of reasoning about the world (such as human others) and we observe it in entities that we believe do not have these faculties (Heider & Simmel, 1944). Arguably, our perceptions of behaviour as intentional are an integral part of the way in which we as humans make sense of the world. They are however not only part of our intuitive understanding; we also theorise entities by ascribing intentions to them. We form explanations and predictions of the behaviour of artificial systems based on the assumption that the behaviour is intentional. This allows us to interact socially with these systems.

Anything, real or virtual, is explicable, explainable and predictable in intentional terms in so far as its behaviour is sufficiently rational or goal-directed. It should be noted that artificial and natural systems are reason-guided and goal-directed to different degrees. Intentionality is sometimes over-attributed to things that are not as reason-guided as they are thought to be. In such cases these things will not behave as expected, which can have detrimental consequences for our understanding of them. The teleological explanations of natural science that were used to explain natural phenomena before the arrival of the mechanistic worldview and the scientific method in the Enlightenment era may in some respect be examples of such over-attributions. Contemporary problems of over-attribution can be found in human-computer interaction in cases where people sometimes expect artificial systems to be ‘smarter’ than they actually are.

While simpler entities can be theorised in physical or mechanistic terms alone, intentionality-ascriptions are practically indispensable for making sense of entities that exhibit higher levels of behavioural complexity and sophistication. Human behaviour is a case in point, as cogently illustrated by Jerry Fodor:

If you want to know where my physical body will be next Thursday, mechanics—our best science of middle-sized objects after
1.2. WHY STUDY ARTIFICIAL AGENT EMBODIMENT?

all, and reputed to be pretty good in its field—is no use to you at all. Far the best way to find out (usually in practice, the only way to find out) is: ask me! (Fodor, 1987, p. 6)

The current state of technology makes it possible to create artificial systems that communicate their own intentions, beliefs and desires in social interaction with humans. There are, for instance, contemporary computer systems that can provide sufficient answers to the above question ‘Where will you be next Thursday?’ – an answer that contains information about where the system intends to be at a particular time. The arrival of the early human writing systems dating several thousand years back represents the genesis of artefacts endowed with the capability to communicate to others the complex intentions of their creators. It is only in the last decades that humans have invented artefacts capable of communicating intentions. It is only very recently that we have begun to see practical need for it in artificial systems. We now see it in the emerging industry of autonomous cars where the driver is replaced by an artificial system. Here we face the problem of how to coordinate the behaviour of autonomous cars with the behaviour of humans, both on the road and at the pedestrian crossing. With the intention of putting robots to use in health care and in domestic contexts comes the need to communicate intentions to make complex goal-directed behaviour comprehensible to those with whom the robots will interact. Agents with complex behaviour that does not make sense to others simply do not interact and coordinate well with others.

We do not need to ascribe intentionality to make sense of the behaviour of many relatively simple entities that are sufficiently described in terms of abstract function or mechanism, such as bacteria, electrical lamps and thermostats. A future scenario where all artificial systems perform socially is implausible since creating socially proficient forks, hammers and stools may provide us with little or no additional value. However, future systems whose complex actions are expected to coordinate well with humans – systems whose whys and wherefores are neither readily understood by human intuition nor comprehensibly explained in terms of mechanism – will have to perform socially.

1.2 Why study artificial agent embodiment?

This section is concerned with two questions: what artificial agent embodiment is and why we should study it. Ziemke (2003, p. 1) noted that
although ‘there are very different notions [within cognitive science] of exactly what embodiment is and what kind of body is required for what type of embodied cognition’ embodiment is ‘by many researchers considered a \textit{conditio sine qua non} for any form of natural or artificial intelligence’. This view reflects a belief in the importance of the notion of embodiment in explaining cognitive phenomena and a concurrent uncertainty as to exactly what embodiment is. Embodied cognition is a relatively young field of research within cognitive science that has grown in popularity in recent years in part due to its success as an empirical field of inquiry, and in particular its demonstrated ability to explain cognitive phenomena with reference to the body of the organism rather than internal mental representation and ‘disembodied’ computation. However, its central construct ‘embodiment’ still defies clear and uncontroversial interpretation. This parallels the state of the central notion of ‘intelligence’ within artificial intelligence research, and, as noted by Woolridge and Jennings (1995, p. 116), the notion of ‘agency’ within agent-based research community.

Regardless of how the notion of embodiment fits into the framework of a general theory of cognition, we can analyse the meaning of the term to sketch a picture of how it is understood and used. The word has at least two different and significant uses. The first one – let’s call this sense of the word embodiment$_1$ – occurs in expressions such as ‘cognition is embodied’ and is used to express the view that having a body is a prerequisite for being a cogniser. Predications of embodiment$_1$ often yield trivial expressions that align with mainstream views on cognition. Embodiment$_1$ is therefore often used to emphasise the importance of the body to cognitive processes rather than to express a particular and distinct theory of cognition.

The word embodiment is also used in a second sense, embodiment$_2$, to predicate the possibility of something either to be disembodied (to have no body) or to have multiple bodies (albeit not necessarily at the same time). embodiment$_2$ occurs in talk about the embodiment of ‘Jesus-on-Earth’ and ‘Jesus-after-Ascension’. Here we talk about an entity, Jesus, that is supposedly able to have multiple bodies and/or no body at all. It also occurs in transhumanist literature in relation to future human minds being \textit{transferred} to artificial bodies or being \textit{uploaded} to a network living in disembodied states of existence$^1$. The reality of the embodiments of Jesus and transhuman minds is supposed by some but is not necessarily true or even metaphysically possible. To the best of our scientific knowledge human beings can only have one (and no less than one) body. If this is knowledge

$^1$For more on mind-body transfer and uploading see Koene, 2012.
1.2. WHY STUDY ARTIFICIAL AGENT EMBODIMENT?

is correct, then it is false to predicate embodiment to humans. However, whether humans can be embodied is an empirical matter which (much like the existence of Descartes’ devious demon or Santa Clause) cannot be decisively negated no matter how implausible the notion seems. The acceptance of the thesis of human embodiment is contingent on evidence to back it up. Although there are views within cognitive science that entail the validity of this thesis the evidence is scarce and inconclusive.

In contrast, there is one predication of embodiment with which we as humans are becoming increasingly familiar, namely: artificial agent embodiment. Artificial agent embodiment refers to the state of either being constructed out of physical materials (robotic embodiment) or appearing to be (but not actually being) constructed out of physical materials (virtual embodiment). An agent of the former type is called a robot and an agent of the latter type is called a virtual agent. There are examples of artificial agents that combine physical and virtual components, such as robots that feature digital displays as faces. It is not clear whether these types of agent bodies represent a third form of embodiment or if they should be classified as robots. Artificial agents are embodied because the behaviour and appearance of an artificial agent can, at large, be realised both in physical and virtual mediums.

Experimental studies comparing robots and virtual agents have shown that robotic and virtual embodiment give rise to different effects on people in social interaction (c.f. Lee, 2015). We now begin to touch upon the second question raised above: ‘Why should we study agent embodiment?’ When confronting the problem of constructing an artificial agent to solve a particular set of problems or to assist in a particular task, one of the most obvious and arguably most important decision is whether to implement a physical or virtual agent. This decision restricts the design space considerably because there is a wide range of things that a robot could do that a virtual agent could not do, and vice versa. In biology morphological variation refers to the differences which exist between members of a species. Variation may be exhibited in physical appearance, behaviour, learning, metabolism, and other characteristics. While morphological variances in humans and other natural agents are relatively small and constant across species and time, artificial agents exhibit a morphological variation that is arguably wider and more extreme than in any existing biological species. Artificial agents range from physically inefficacious virtual agents to strong, fast, precise and tire-

\[2\] A popular way to create this ‘faux state of embodiment’ is by using computer generated graphics and sound.
less physical robots, and from immutable and stationary robots to morphologically flexible and electronically mobile virtual agents. Different agent attributes – some of which are restricted only to robotic or virtual embodiment – lend different agents unique qualities in social interaction. Some of these qualities, such as likability, trustworthiness, and persuasiveness, are largely associated with satisfactory social interaction while others are considered to be detrimental. The success of the agent in fulfilling its purpose is contingent on the designer's knowledge of what qualities and effects various agent attributes give rise to in interaction. This is the primary reason why to study artificial agent embodiment.

1.3 Social influence and the artificial agent

The field of social psychology was defined by one of its pioneers, Gordon Allport, as:

The scientific attempt to explain how the thoughts, feelings and behaviours of individuals are influenced by the actual, imagined, or implied presence or others (Allport, 1954).

According to this definition, social psychology is essentially the study of social influence. As noted by Fiske, ‘social influence affects not only trivial behaviour, such as making paper airplanes in class. It also affects important behaviour, such as yielding to majority opinion over one's own judgement; ... allowing oneself to be imprisoned doing useless paperwork; or torturing innocent people, as soldiers sometimes have done’ (Fiske, 2009, p. 4).

Social psychological theories often seek to explain why people behave as they do in social contexts. A straightforward way to do this is to ask people what their reasons are for behaving in particular ways. However, although people can usually produce an explanation for their behaviour, their explanations are not necessarily true or accurate. This is because the best explanation is not always readily available to an individual either intuitively or by introspection. To illustrate this point, consider the famous 1976 experiment by Nisbett and Wilson, in which experimenters set up a stall in a commercial establishment asking people passing by to evaluate the quality of four identical pairs of socks placed in a row on a table (Nisbett & Wilson, 1977). Although the socks were identical, people tended to prefer the socks positioned at the right of the table with a factor of almost four-to-one. When asked about the reasons for why they preferred the socks to the right, ‘no subject ever mentioned spontaneously the position of the article in the
array. And, when asked directly about a possible effect of the position of the article, virtually all subjects denied it ... (ibid., 1977)’. Social psychological research cannot rely fully on people’s own explanations of their attitudes and behaviour, since people’s self-reports do not always make up the whole picture of what causes them to think or behave in different ways. Thus, an important task of social psychology is to identify the ways in which others exert social influence on the attitudes and behaviour on people.

The term ‘others’ in Allport’s above definition refers to fellow human beings, and it is the thoughts, feelings and behaviours of human others that have been the focus of the vast majority of studies within social psychology to date. However, the emergence of artificial social agents with unique and morphologically flexible embodiments paves the way for profoundly different types of social interactions and dynamics. This provokes the question to what extent existing social psychological theory is applicable to cases of social interaction which involve artificial social agents, and whether its object of study should be expanded to incorporate not only the natural human agent but also the artificial agent. The significance of understanding how artificial agents affect us psychologically in social interaction depends, of course, on the extent to which people actually interact with artificial agents. People all over the world already interact with artificial agents on a daily basis. For instance, intelligent virtual assistants such as Apple’s Siri, Google’s Google Now, and Amazon Alexa are used by millions of people. These agent-based systems exert social influence on us affecting our attitudes, knowledge, and behaviour. As research has shown, we not only make decisions based on the information that these systems provide us with, we are also biased by the nature of the agent and how we perceive of it. For example, Nass et al., (1997) showed that synthesised speech is assigned a gender by listeners, and that such assignments can cue social stereotypes that influence user behaviour. A 1996 study showed that participants of both genders tended to reveal more information to a female voice than a male voice (Tannen, 1996). This creates an incentive for research into the social influence of different types of artificial agents with different embodiments, appearances and behaviours.

1.4 Purpose and structure

The general purpose of the present thesis is to investigate the relationship between agent embodiment, presence and influence. This is done by experimentally assessing the effect of robotic versus virtual embodiment on agent
presence and influence in the context of a decision task. Specific research questions are presented subsequent to reviewing previous studies that motivate them.
Chapter 2

BACKGROUND

2.1 The nature of artificial agents

In his 1969 book *The Sciences of the Artificial*, Herbert Simon proposed a clear-cut dichotomy between the natural sciences and ‘the sciences of the artificial’. While the natural sciences describe objects and phenomena that occur naturally in the world, the sciences of the artificial describe objects and phenomena that result from human intervention in the world, i.e. ‘artefacts’. According to Simon, artefacts are fundamentally different from natural objects in that their form arises from the human purposes they are built to serve. He proposed that the proper study object for the sciences of the artificial is the abstract functional or ‘purposeful’ aspect of artificial things. Simon saw the purposes of an artefact as a link between its internal physical structure and its environment, and he held that we can study the function of the artefact largely without studying its internal structure or environment.

It is easy to see why one would think that artefacts can be studied with reference to its function alone, especially when considering how artefacts are normally conceptualised and described. Artefacts are described in terms of *abstract functional specification*. Such specifications come in the form of written-down-words or a combination of words and imagery (so called ‘schematic diagrams’). Observe the circuit diagram of an electric lamp in Figure 2.1. The diagram contains information about the purpose or function of the lamp (how it functions to produce light) in terms of functional components such as a ‘power source’ and an ‘electrical switch’. This information can be viewed as a set of requirements that must be fulfilled in order for something to be an electrical lamp. Note that a functional specification of a lamp says nothing in particular about what materials a lamp is made out
of; it leaves open the logical possibility that the lamp can be made out of any physical material as long as it meets the functional requirements. We know, for instance, that the electrical wiring of an electrical lamp can be made out of silver, copper or aluminium, and that tungsten, carbon or osmium can be used as filament. Many artefacts are, in this sense, to quite a large extent multiply realisable. However, they are not realisable in just any substrate. There are a lot of different materials that we know cannot be used to build an electrical lamp. Lamps do not come with stone or spaghetti filament or with rubber wiring, for instance. Arguably, the same goes for any artefact: the extent to which it is multiply realisable is determined by the variety of physical materials that can sustain its function(s).

The purposes that artificial systems are built to serve are unexceptionably central to the study of these systems. However, the nature of artificial systems cannot be explained in terms of purposes alone, since what purposes can be fulfilled by a particular system is constrained by its physical nature. The nature of artificial social agents, such as social robots and virtual agents, cannot be described with reference only to the purposes that they fulfil in their respective operational domains; we also need knowledge of how different agent embodiments and morphologies constrain the agent’s functionality and the purposes that the agent ultimately can fulfil in social interaction. Bill Hillier noted that the ‘inner physical structure’ of artefacts, the study of which Simon viewed as inessential to the study of the artificial, is in fact a common study object where ‘artificial systems are found with the curious property of being both man-made and not understood by man’ (Hillier, 1985, p. 164). Hillier counts language, cities and societies to the list of such systems. The same is undoubtedly true of artificial agents – despite having created them we know remarkably little about what makes them purposeful, functional or, indeed, social. If we want to create purpose-
2.1. THE NATURE OF ARTIFICIAL AGENTS

ful social agents we have to study different types of agent embodiments and morphologies as well as how they perform in social interaction.

Several attempts have been made to provide a definition of ‘agent’. In a 1995 paper discussing the term, Woolridge and Jennings referred to a remark originally made by Carl Hewitt in 1994 that ‘the question “what is an agent?” is embarrassing for the agent-based computing community in just the same way as the question “what is intelligence?” is embarrassing for the mainstream AI community’ (Woolridge & Jennings, 1995, p. 116). Definitions proposed by different authors reflect different view on what is to be expected by a definition of the term. Several authors have proposed definitions that attempt to state necessary and sufficient conditions for when something is an agent. For example, Franklin and Graesser proposed that:

An autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future. (Franklin and Graesser, 2005, p. 25)

However, such definitions tend to run into boundary problems. As noted by the authors themselves the above definition fails to exclude things that are generally not considered to be agents, such as thermostats, fire detectors or bacteria. In response to this, some authors have chosen to adopt an instrumentalist position on the matter of how to define the term. For example, AI scientists Russell and Norvig broadly defined agent as ‘anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors’, arguing that ‘the notion of an agent is meant to be a tool for analyzing systems, not an absolute characterisation that divides the world into agents and non-agents’ (Russell & Norvig, 1995, p. 33). Franklin and Graesser questioned the usefulness of instrumentalist definitions, such as Russell and Norvig’s, stating that:

If we define the environment as whatever provides input and receives output, and take receiving input to be sensing and producing output to be acting, every program is an agent. Thus, if we want to arrive at a useful contrast between agent and program, we must restrict at least some of the notions of environment, sensing and acting. (Franklin & Graesser, 2005, p. 22, my italics)

Another proposed definition of agent is that agents are ‘intentional systems’. Daniel Dennett defined an intentional system as ‘a system whose
behavior can be (at least sometimes) explained and predicted by relying on ascriptions to the system of beliefs and desires (and hopes, fears, intentions, hunches, ...) (Dennett, 1971, p. 87). In other words, according to this viewpoint the necessary and sufficient condition for something to be an intentional system, and therefore an agent, is that an external observer can explain and predict its behaviour by relying on ascribing beliefs, desires and similar folk psychological entities to it.

The fact that individuals can erroneously ascribe intentions to things, by ascribing intentionality to something that does not behave in accordance with the assumption that it possesses intentionality, is a problem for this type of theory. The theory that an agent is an intentional system represents a radically different attempt to define the notion. Arguably, this attempt avoids the boundary problem of delimiting agents from non-agents but at the cost of altogether blurring the boundary. There seems to be no definite way to settle the matter of whether a particular entity, say a robot, is an agent or not by relying on people’s error-prone intuitions about what or who has beliefs and desires. Rather it seems that the attempt pushes the problematic aspects of the notion ‘agent’ over to the equally problematic notion ‘intentional system’.

In conclusion, not much progress has been made in the search for a definition of ‘agent’ that meets scientific standards. The term is slippery and affected by boundary problems. It is also unclear what to expect from a definition of it. The author of this thesis adopts a position of suspended judgement with regard to the question of what an agent really is, and chooses to focus on general rather than ‘essential’ characteristics and qualities of artificial agents.

## 2.2 The making of social agents

The concepts ‘agent’ and ‘social’ are clearly interrelated. Being social seems to imply also being an agent. However, being social does not necessarily follow from being an agent. Being social can be described as being positioned somewhere between the two end-points of a continuum, where one extreme represents ‘highly social’ and the other ‘not social at all’. An agent’s position on this continuum is contingent upon behaving in certain ways considered social. What constitutes ‘social behaviour’, and how do we approach the problem of building artificial agents that exhibit it? Carley and Newell suggested that we put a spin on the classic Turing test:

**The Social Turing test:** Construct a collection of artificial social
agents according to the hypotheses about what makes agents social and put them in a social situation, as defined by the hypotheses. Then recognizably social behavior should result. Aspects not specified by the hypotheses, of which there will be many, can be determined at will. The behavior of the system can vary widely with such specification, but it should remain recognizably social. (Carley & Newell, 1994, p. 257)

The Social Turing test, just like the classical Turing test, relies on the recollection of an observer to test for sufficiency (in this case social sufficiency). It describes a process of hypothesis testing in which we as creators are free to test any hypotheses that we have about what properties agents need to have to become sufficiently social. In Carley and Newell’s words, the Social Turing Test ‘tests the proposition that, if the agent has properties x, y and z, then it behaves socially, on the assumption that humans can recognize social behavior in all of its forms’ (ibid., p. 257).

The Social Turing test is based on the assumption that what counts as social behaviour is relative to the behavioural norms and expectations within particular social communities. Carley and Newell point out that being social is not primarily about what one can do; rather it is about acting under the constraints of a normative system:

Socialness arises not from capabilities but from limitations. Socialness is a response to environmental complexity (the presence of multiple others, multiple and simultaneous goals, rich cultural-historic heritage, and so on). (ibid., p. 260)

Being subjected to social norms that result from the multiplicity of other’s intentions, goals, desires, and so on restrict our actions and tune us to produce socially appropriate responses. Social behaviour is, in this sense, not a product of deliberate internal reflection but of environmental pressure imposed on agents without them being consciously aware of it. Despite seldom being aware of the extent to which our behaviour is determined by norms, or being able to put the nature of these norms to words, people are quite good at detecting deviations from them. Deviant or unwanted behaviour provoke attitudinal and behavioural responses that can be observed and measured. This makes identifying social behaviour largely a matter of ‘knowing when seeing it’, or better yet: ‘knowing when not seeing it’. This is also what makes the Social Turing Test work, and what legitimises making inferences about the social qualities of agents based on people’s interactions with them.

---

This paper was published two years after the death of Allen Newell

---

1
Social environments are not only complex but also largely different from each other. As pointed out by Fong, Nourbakhsh, and Dautenhahn social qualities are not in themselves ‘good’ or ‘bad’ but are valued differently across different environments:

Social behavior includes such a wide range of phenomena that it is not evident which features a robot must have in order to show social awareness or intelligence. Clearly, a robot’s design depends on its intended use, the complexity of the social environment and the sophistication of the interaction. (Fong, Nourbakhsh & Dautenhahn, 2003, p. 160)

To illustrate this point, consider a social agent with the quality of being talkative and outgoing. This quality may in many situations facilitate social interaction, but in others it could be detrimental. For example, a chatty agent in a car may distract the driver from the road – or perhaps, as suggested by Eriksson and Stanton (2016), it could be used as a co-driver helping the driver of a semi-autonomous vehicle to keep his or her attention on the road. Moreover, an agent that gives off the impression that it is reliable and trustworthy can be useful and appropriate in cases where the performance of the agent measures up to people’s expectations on it, but outright dangerous in cases where it does not.

2.2.1 Design and creation

Given that the many social qualities are considered beneficial in some environments and detrimental in others, the domain in which an agent is intended to operate should be taken as the starting point for its making. The first step in the creation process is to identify a set of hypothetical tasks and events that the agent is probable to confront in its intended domain of operation. We then specify what types of behavioural responses are appropriate or required given the set of hypothetical events previously formulated. We can define these kinds of ‘mappings’ between behavioural responses (valid solutions) and tasks (problem) as task requirements. All the specified task requirements of a domain of operation taken together may be called a task domain specification:

**Task domain specification:** a mapping of agent responses and tasks guided by the question ‘What tasks will the agent be required to carry out?’
When the task domain of the agent has been specified, the next issue is what social qualities the agent needs in order to function socially in the task domain. For example, assuming that we have a task domain mainly focused on assisting people when driving a semi-autonomous car, a task requirement may be for the agent to issue a ‘take-over-request’ in case the self-driving system of the car cannot handle the road situation. An appropriate social quality for this agent may be to be commanding or self-confident. In contrast, a robots used in health care that are required to interact with very sick people perhaps need to be supportive and consolatory. We will call the specification of required agent qualities given a particular task domain specification a quality specification:

**Quality specification:** a mapping of tasks to agent qualities, guided by the question ‘What agent qualities will satisfy the requirements within the task domain?’

After quality specification comes the task of specifying what attributes the agent needs in order to realise the desired qualities in social interaction. Here we face questions regarding the agent’s embodiment, appearance and behaviour, such as: ‘Is there a need for a physical body?’, ‘Is non-verbal communication, such as gestures, important?’, ‘Should the agent be human-like and, if so, to what extent?’, and ‘Should the agent be endowed with gender?’. Qualities mapped to attributes result in a morphology specification:

**Morphology specification:** a mapping of agent qualities to agent attributes, guided by the question ‘What agent attributes will give rise to the desired qualities?’

The last step of the design process is to identify the computational and material requirements needed to build the agent and create a material specification:

**Material specification:** a mapping of agent attributes to physical or computational requirements, guided by the question ‘How do we implement the specified agent attributes?’

Here we face questions such as ‘How do we successfully implement anthropomorphic appearance and behaviour in an artificial agent?’ and ‘How do we build a robot that is physically shaped so as to support people with walking disabilities?’ See Figure 2.2 for an overview of the design process.
2.2.2 Evaluation and ideation

Continuous, systematic evaluation is required for artificial social agents to improve. As suggested here the evaluation process should be focused on two overarching purposes:

- **Validation of social sufficiency**
- **Investigation of morphological causes for social sufficiency**

Validation of social sufficiency is the process of assessing whether or not, or to what degree, the agent possesses the social qualities that it is designed to have. It answers the question, ‘Is the agent (sufficiently) social?’ This is essentially the process described above as a Social Turing Test, where we rely on people’s intuitions to test for social behaviour. Investigating the morphological causes for social sufficiency means answering the question, ‘In virtue of what agent attributes is the agent socially sufficient?’ This question needs to be assessed using a experimental or quasi-experimental setup.

**Validation of social sufficiency**

There are practical benefits to having a formal quality specification at hand when evaluating the social sufficiency of the agent. A quality specification
can function as a guide to which social qualities of the agent should be tested during evaluation. For example, if an agent according to its quality specification is intended to be authoritative, trustworthy and a good teacher we focus primarily on assessing these qualities. If an agent is specified to be likeable, life-like and persuasive we focus on those qualities. It is also a good idea to measure constructs that may indicate overall social sufficiency, such as ‘social presence’, for the purposes of comparing different artificial agents with each other.

The measures that are used to measure social qualities of the agent can be roughly grouped into two categories: *attitudinal measures* and *behavioural measures* (see Table 2.2). Attitudinal measures are those that measure (psychological) effects on people’s attitudes elicited from interacting with an agent. One of the most common types of attitudinal measures is the post-test self-report questionnaire, where people are asked to rate their experience from interacting with an agent with regard to some quality on part of the agent. Self-report questionnaires that can be used pre-test to target participant biases are also common. Behavioural measures target effects on people’s behaviour that arise when interacting with an agent. These effects are measured using a variety of techniques including: logging data points to measure an individual’s performance in a given task, paper note-taking of how many right answers participants’ score in a quiz or a recollection task, and measuring symptoms of arousal such as increased pulse.

<table>
<thead>
<tr>
<th>Attitudinal measures</th>
<th>Behavioural measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likeability</td>
<td>Persuasiveness</td>
</tr>
<tr>
<td>Appeal</td>
<td>Ability to teach</td>
</tr>
<tr>
<td>Trustworthiness</td>
<td>Ability to attract and direct attention</td>
</tr>
<tr>
<td>Reliability</td>
<td>Ability to arouse</td>
</tr>
<tr>
<td>Social presence</td>
<td>Ability to inspire prosocial behaviour</td>
</tr>
<tr>
<td>Life-likeness</td>
<td>Ability to collaborate</td>
</tr>
<tr>
<td>etc...</td>
<td>etc...</td>
</tr>
</tbody>
</table>

**Investigating the morphological causes of social sufficiency**

Assessing the question *why* or *in virtue of what* an agent is social means inferring causes related to the embodiment and morphology of the agent. In order to make such inferences some attribute(s) of the agent must be experimentally manipulated while other agent attributes are held constant. See
Table 2.3 for examples of agent attributes.

<table>
<thead>
<tr>
<th>Embodiment</th>
<th>Behaviour</th>
<th>Appearance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical vs virtual body</td>
<td>Autonomy</td>
<td>Anthropomorphism</td>
</tr>
<tr>
<td>Physical vs virtual presence</td>
<td>Gestures</td>
<td>Zoomorphism</td>
</tr>
<tr>
<td></td>
<td>Gaze behaviour</td>
<td>Caricature</td>
</tr>
<tr>
<td></td>
<td>Learning</td>
<td>Functional</td>
</tr>
<tr>
<td></td>
<td>Speech</td>
<td>Gender</td>
</tr>
<tr>
<td></td>
<td>etc...</td>
<td>etc...</td>
</tr>
</tbody>
</table>

*e.g. low, moderate or high levels of autonomy*

**Ideation**

The issue of how to design a particular artificial agent is largely determined by the requirements of the tasks to be performed, and the success of the agent in fulfilling its purpose is contingent on the designer’s understanding of the task domain. While many tasks are well defined and have clear goals, such as sorting out and removing components of type x from an industrial production line, tasks that are characterised as ‘social’ often lack a clear solution. Such tasks may involve acting such as not to infringe on the social integrity of others, being able to predict and respond to the actions of others proactively, or simply ‘being nice’ to others. When writing a quality specification for the agent the designer in effect formulates an hypothesis about which social qualities are relevant or necessary for the agent to successfully meet its purpose. This process of hypothesis formulation needs to be informed by results from studies that have validated the social sufficiency of different agents in different contexts. We can think about the formulation of an agent morphology specification in a similar way as a process of ideation which needs to be substantiated by results from research investigating the morphological causes of social sufficiency.

We need well informed, empirically supported ideas to make well functioning agents. We need to be conscious about how to research human-agent social interaction so that we can better integrate the results from such research into the design process. The people who design and create artificial agents are often not the same as the people who evaluate and in-
2.2. THE MAKING OF SOCIAL AGENTS

Knowledge from research into agent-based social interaction, such as from human-robot interaction and intelligent virtual-agent research, needs to be better integrated into the design and creation of new artificial agents. This requires a systematic effort from the research community, where validations of social performance and findings on the connection between social qualities and different agent morphologies are made sensible to designers. At the same time, researchers evaluating human-agent interaction would arguably benefit from a comprehensive specification of task domains, qualities and attributes of the artificial agents that are already ‘out there’ and available for evaluation. This thesis argues for a tighter coupling between on the one hand the design and creation process and on the other hand evaluative and investigative research. It argues that the way to achieve this ‘tighter coupling’ is through comprehensive formal specification of artificial agent properties and through applying the experimental method to assess social sufficiency and to investigate its morphological causes (see Figure 2.3). This way of doing artificial agent research will in the long run bring about better hypotheses about what causes artificial agents to perform socially in the context of particular tasks.

Figure 2.3: The reciprocal relationship between agent design and evaluation
2.3 Agent embodiment and presence: previous findings

Li (2015) made a comprehensive survey identifying 33 key works comparing physical and virtual agents. These studies ranged in publication year from 2001 to 2013, included wide variety of agents, task environments, study populations, and employed a variety of different measures to examine effects on various aspects of social human-agent interaction.

Li conducted a meta-analysis based on his survey which showed that, out of 62 significant effects found in the studies covered in the survey, 73% favoured robots over virtual agents, 21% favoured virtual agents, and 6% were crossover interaction effects that varied depending on participant age, task type, and presence of robot gestures. The outcome of this meta-analysis was thus that robots in general elicit more favourable (both attitudinal and behavioural) responses than virtual agents in social interaction contexts.

Moreover, Li found that participants only favoured robots over virtual agents when they interacted with robots that were located in the same room as themselves. They did, however, not prefer robots over virtual agents when robots were presented to them on a computer screen. According to Li, this suggests that the physical presence of robots has a larger impact on user response than their physical embodiment. Based on this observation Li made a methodological recommendation. Co-located robots are physically embodied as well as physically present. In contrast, virtual agents are virtually embodied and virtually present. Consequently, when the two types of agents are evaluated in comparison with each other there is a risk that effects of (physical versus virtual) agent embodiment are conflated with effects of (physical versus virtual) presence. Li recommends that the two different attributes should be investigated separately: if one is interested in examining effects of embodiment one should, according to Li, compare tele-present robots with virtual agents (agent presence is held constant); if one is interested in the effects of presence one should compare co-located robots with tele-present robots (agent embodiment is held constant).

The present section of the thesis highlights a few experimental results that are of particular interest to the purposes of investigating the effects of agent embodiment and presence on agents’ social influence on peoples’ behaviour and decision making, which is the focus of the study presented in this thesis.
2.3. AGENT EMBODIMENT AND PRESENCE: PREVIOUS FINDINGS

Effects of agent embodiment and presence on human behaviour

Kiesler et al. (2008) conducted a study which manipulated embodiment (physical versus virtual agent) and presence type (present versus projected agent). The presence types were defined as follows: a ‘present’ physical agent is a co-located robot, a ‘present’ virtual agent is an agent presented to the user on a computer screen, a ‘projected’ physical agent is a video-feed of a robot projected onto a wall, a ‘projected’ virtual agent is an agent projected onto a wall. Study participants were told that their goal was to discuss basic health habits with the agent (including topics like exercise, diet, weight and height, and teeth flossing). During the conversation the participants were offered candy bars or health bars that were placed on a table in front of them. The social influence of the agents was measured using two response variables related to eating: ‘Did the participant eat health bars rather than candy bars?’ and ‘Did the participant eat fewer calories?’.

The results showed that the presence type of the agent influenced the participants’ responses to the offer of snacks. Participants that interacted with a present agent were more likely to choose a health bar than a candy bar in comparison to those who interacted with a projected agent. Furthermore, those who interacted with the present robot ate fewer calories than those who interacted with the projected (tele-present) robot. The conclusion was that participants behaviour was subconsciously influenced by agent type, with present ‘agents’ exerting greater influence than ‘projected’ agents.

In a 2010 study by Bainbridge et al., participants were presented with a book sorting task where they were asked to follow the instructions of an artificial agent indicating in which bookshelf a particular book should be placed. The participants were divided into three groups and were subjected to one out of three experimental conditions featuring different artificial agents. The conditions were: a ‘physical condition’ with a co-present physical agent (i.e., a robot), a ‘live-video condition’ with the same robot presented on a computer screen (i.e., a telepresent robot), and an ‘augmented-video condition’ with the telepresent robot as in the previous condition but with two computer screens which presented the robot from two different angles. The last condition was intended to ‘balance the loss of three-dimensional information in the live-video condition’ (Bainbridge et al., 2010, p. 43). At one point in the experiment, the participants were asked to cooperate in an ‘unusual’ or ‘destructive’ task. The robot pointed to a pile of expensive-looking textbooks and then to a garbage can placed in the room. Participants in all three conditions expressed hesitation or confusion at the request to put the books in the garbage can. However, they chose different courses of ac-
tion depending on what condition they were in. A statistically significantly higher proportion of participants in the physical condition (12 out of 20) followed the instruction and threw away the books as requested by the agent than in the live-video condition (2 out of 18) and augmented-video condition (3 out of 18). From this result it can be concluded that the physical agent exerted greater social influence on the participants in comparison to the agent displayed on screen. However, as noted by the authors it is not obvious exactly what quality of the physical agent gave rise to increased social influence. They speculate that:

This combination of interactive behaviour, and post-interaction, self-reported perceptions, indicates that participants afford greater trust to the physically present than to the video-displayed robot, making participants more willing to follow through with an unusual request from the robot. On the other hand, this could instead indicate that physical presence increases a participant's desire to comply with her social partner. (ibid., p. 50)

### 2.4 Research questions

This thesis answers questions regarding the embodiment and social presence of artificial agents that have been formulated based on previous findings within the agent-based computing research field. It aims to shed light on how these aspects of artificial agents affect their social influence on people's decisions, including decisions to behave prosocially in ways that intentionally benefit another. It also aims to contribute to the understanding of how the negativity of people's attitudes toward artificial agents relate to how they are perceived, in particular to whether they are perceived of as social actors or merely machines.

**Embodiment**

Three hypotheses regarding the effect of embodiment on people's decisions and perception of agents were formulated based on previous research on the benefits of physical agents over virtual agents in social interaction with people (see Li, 2015):

**First hypothesis regarding embodiment (EMB1):** Physical embodiment gives rise to a higher degree of social influence than virtual embodiment.
Second hypothesis regarding embodiment (EMB2): Physical embodiment elicits a higher degree of prosocial behaviour in individuals interacting with the agent when compared to virtual embodiment.

Third hypothesis regarding embodiment (EMB3): Physical embodiment gives rise to perceptions of the agent as socially present to a higher degree than virtual embodiment.

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Social influence</th>
<th>Prosocial behaviour</th>
<th>Social presence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical EMB</td>
<td>EMB1</td>
<td>EMB2</td>
<td>EMB3</td>
</tr>
<tr>
<td>Virtual EMB</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Social presence

Based on the previous research highlighting the benefits of artificial agents being perceived of as highly socially present, two conjectures were made regarding the effect of agent social presence on people's decision making:

First hypothesis regarding social presence (SP1): The agent is more socially influential when perceived as highly socially present (than when perceived as socially present to a lesser degree)

Second hypothesis regarding social presence (SP2): The agent elicits higher degrees of prosocial behaviour in individuals interacting with it when perceived as highly socially present (than when perceived as socially present to a lesser degree).

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Social influence</th>
<th>Prosocial behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td>High SP</td>
<td>SP1</td>
<td>SP2</td>
</tr>
<tr>
<td>Low SP</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Negative attitude and social presence

People’s negative attitudes toward artificial agents have been studied extensively. However, the relationship between a person's negative attitude
toward a particular artificial agent and the extent to which he or she perceives of the agent as a socially present actor remains unestablished. The following hypothesis is based on the assumption that the tendency of people not to perceive of artificial agents as social actors may to some extent be caused by negative attitudes toward artificial agents (although this assumption is not directly assessed by the testing of the hypothesis, it is a first step toward linking together two possibly related aspects of artificial agent social interaction):

**Hypothesis regarding negative attitude and social presence (NASP):** There is a relation between people's negative attitudes toward artificial agents and the extent to which they perceive of them as socially present, such that high levels of negativity toward artificial agents positively correlate with low levels of perceived social presence and *vice versa*.

<table>
<thead>
<tr>
<th>Correlate variables</th>
<th>Negative attitude</th>
<th>Social presence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative attitude</td>
<td>-</td>
<td>NASP</td>
</tr>
<tr>
<td>Social presence</td>
<td>NASP</td>
<td>-</td>
</tr>
</tbody>
</table>

**Gender effects**

Gender effects have been studied before in the context of the human-to-human ultimatum game paradigm (which will be presented in the following section). However, the effect of gender has to the authors knowledge not yet been studied in the context of a human-robot ultimatum game. Gender effects will therefore be reported in the results section of the study report.
Chapter 3

METHOD

3.1 Design

The proposed hypotheses demanded establishing both correlation (NVSP) and relations of cause-and-effect (EMB1–3, SP1–2). Specifically, inferences had to be made about the effects of particular agent attributes on particular agent qualities. Because of this, a controlled experimental setup was chosen as the method by which to assess the hypotheses. The experimental design featured a between-group element with agent attribute ‘embodiment’ as the manipulated factor, that is, with one group of participants tasked to interact with a physical agent (robot) and a second group with participants tasked to interact with a virtual agent. Agent attributes other than the manipulated factor (embodiment) were, to the largest extent possible, held constant in order not to confound presumptive observations of effects caused by the manipulated factor. To this end the Aldebaran Nao agent platform, which features a NAO-robot and a near identical virtual representation of the NAO-robot, was utilised (see Figure 3.1).

3.2 Instruments

3.2.1 Social influence and prosocial behaviour

Social influence was measured based on the premise that the degree of social influence of the agent on participant decision making is reflected in the participants’ inclination to accept an offer made by the agent in an ultimatum game. The rules of the ultimatum game is as follows:
1. An amount of money is to be distributed between two players.

2. One of the players, the ‘proposer’, is given the task to propose a freely chosen proportion of the money to other player, the ‘responder’. 

3. The responder is given the task to either accept or decline the offer made by the proposer.

4. The consequence of accepting the offer is that both the proposer and the responder will be compensated in accordance with the proposer’s proposition.

5. The consequence of declining the offer is that neither the proposer or the responder will receive any money.

In the experiment participants were given the role of the respondent and the robotic or virtual agent was given the role of the proposer. Participants were offered a fixed amount of 20 SEK out of a pool of 100 SEK (at the time of writing approximately €11). The decision to fix the offer at the ratio of 20 percent was based on a study by Nitsch and Glassen (2015) in which participants tended to decline offers below 20 percent and accept offers above 20 percent. If the fixed amount to offer would be higher, for instance 40 percent, there is an increased risk of committing a type II error and to fail to detect an effect that is present, due to the presumably small number of people that would decline the offer. Similarly, if the fixed amount would have been lower there is an increased risk to commit a type II error due to the small number that would accept such an offer.

The social influence of the agent on participants’ inclination to behave prosocially was measured using two different measures: the number of voluntarily corrected math statements made by the agent, and the decision to
donate or not to donate the acquired money back to the agent. Regarding the former measure, the participants were asked by the agent to help it with learning mathematics by correcting it when stating answers to math problems. The problems were based on the 14 times multiplication table and the statements were given verbally by the agent on the form ‘Four times six is twentyfour’, ‘Five times fourteen is seventy’, and so on. The exercises were given in sets of three, and after each set participants were given the choice to continue or to quit helping the agent. The second measure was whether the participant donated money back to the agent. Only participants that accepted the offer made by the agent were confronted with this decision at the end of the session. The number of corrected math statements were both taken as indicators of degree of prosocial behaviour.

All participant decisions – accept or decline offers, number of math exercises and decision to donate acquired money – were measured using an input device (Figure 3.2) and logged by the computer which run the agent behaviour. The input device was constructed specifically for the purpose of letting the participants respond by giving non-verbal ‘yes’ (green button) or ‘no’ (red button) answers. The primary reason why not to use verbal communication from the participants to the agent was that it was difficult to construct a speech interface for the virtual agent which functioned sufficiently similar to that of the robot.

Figure 3.2: Input device

3.2.2 Negative attitude scale

The level of negativity of participants’ attitudes toward artificial agents was measured using the Negative Attitudes toward Robots Scale (NARS; Nomura et al., 2006). NARS has been used previously in a range of scenarios to identify several factors which affect people’s negativity toward robots, including gender, age, prior experience with robots and cultural differences.
The scale includes 14 questionnaire items divided into three subordinate scales (Table 3.1), covering different kinds of negative attitudes (Table 3.2). General negativity can be measured by combining the scales. Each questionnaire item is a statement which the respondent assigns a value on a scale of zero to five depending the extent to which he or she agrees with it. The value of zero represents strong disagreement with the item statement (and a minimum level of negativity) and the value of five represents strong agreement with the item statement (and a maximum level of negativity). The scale was translated from English to Swedish.

### Table 3.1

**NARS subordinate scales**

<table>
<thead>
<tr>
<th>No.</th>
<th>Subscale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Negative attitudes toward situations of interaction with robots</td>
</tr>
<tr>
<td>2.</td>
<td>Negative attitudes toward the social influence of robots</td>
</tr>
<tr>
<td>3.</td>
<td>Negative attitudes toward emotions in interaction with robots</td>
</tr>
</tbody>
</table>

### Table 3.2

**NARS items**

<table>
<thead>
<tr>
<th>No.</th>
<th>Content</th>
<th>Subscale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>I would feel uneasy if robots really had emotions</td>
<td>S2</td>
</tr>
<tr>
<td>2.</td>
<td>Something bad might happen if robots developed into living beings</td>
<td>S2</td>
</tr>
<tr>
<td>3.</td>
<td>I would feel relaxed talking with robots</td>
<td>S3</td>
</tr>
<tr>
<td>4.</td>
<td>I would feel uneasy if I was given a job where I had to use robots</td>
<td>S1</td>
</tr>
<tr>
<td>5.</td>
<td>If robots had emotions, I would be able to make friends with them</td>
<td>S3</td>
</tr>
<tr>
<td>6.</td>
<td>I feel comforted being with robots that have emotions</td>
<td>S3</td>
</tr>
<tr>
<td>7.</td>
<td>The word “robot” means nothing to me</td>
<td>S1</td>
</tr>
<tr>
<td>8.</td>
<td>I would feel nervous operating a robot in front of other people</td>
<td>S1</td>
</tr>
<tr>
<td>9.</td>
<td>I would hate the idea that robots or artificial intelligences were making judgments about things</td>
<td>S1</td>
</tr>
<tr>
<td>10.</td>
<td>I would feel very nervous just standing in front of a robot</td>
<td>S1</td>
</tr>
<tr>
<td>11.</td>
<td>I feel that if I depend on robots too much, something bad might happen</td>
<td>S2</td>
</tr>
<tr>
<td>12.</td>
<td>I would feel paranoid talking with a robot</td>
<td>S1</td>
</tr>
<tr>
<td>13.</td>
<td>I am concerned that robots would be a bad influence on children</td>
<td>S2</td>
</tr>
<tr>
<td>14.</td>
<td>I feel that in the future society will be dominated by robots</td>
<td>S2</td>
</tr>
</tbody>
</table>
3.2.3 Social presence scale

A scale adopted from Lee et al. (2006) was employed to measure the underlying construct ‘social presence’. The scale consisted of eight questions translated into Swedish and had a high level of internal consistency, as determined by a Chronbach’s $\alpha$ of 0.895. The first three questions were answered on a 10-point semantic differential scale and instructed the participant to ‘for each pair of adjectives place a cross at the point between them which reflects the extent to which you believe the adjectives describe the robot’. The adjectives are presented in Table 3.3.

<table>
<thead>
<tr>
<th>No.</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Unsociable / Sociable</td>
</tr>
<tr>
<td>2.</td>
<td>Machine-like / Life-like</td>
</tr>
<tr>
<td>3.</td>
<td>Insensitive / Sensitive</td>
</tr>
</tbody>
</table>

The following five questions were answered on an independent 10-point Likert scale (0: not at all, 10: completely):

<table>
<thead>
<tr>
<th>No.</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.</td>
<td>While you were interacting with this robot, how much did you feel as if it were an intelligent being?</td>
</tr>
<tr>
<td>5.</td>
<td>While you were interacting with this robot, how much did you fell as if it were a social being?</td>
</tr>
<tr>
<td>6.</td>
<td>While you were interacting with this robot, how much did you feel as if it were communicating with you?</td>
</tr>
<tr>
<td>7.</td>
<td>While you were interacting with this robot, how much attention did you pay to it?</td>
</tr>
<tr>
<td>8.</td>
<td>While you were interacting with this robot, how much did you feel involved with it?</td>
</tr>
</tbody>
</table>

3.3 Participants

60 Swedish university students (40 male, 20 female; median age of 23) took part in the study. A majority were computer science students. Participants were selected with a convenience sampling strategy, and were randomly assigned to one of the two experimental conditions.
3.4 Procedure

Prior to the experiment each participant was informed about the procedure, treatment of experimental data, their rights as participants to abort the experiment at any time, and that there was an opportunity to receive economic compensation depending on actions taken in the experiment. All participants gave their written consent.

Participants were then instructed to take a seat in front of the agent, which was situated on a table approximately one meter away from the participant, and the interaction session was started. The agent first explained the rules of the ultimatum game: (1) that there is a pool of money, in this case an amount of 100 SEK, that is to be distributed between the agent and the participant; (2) that it is up to the agent how to split the money with the participant; (3) that it is up to the participant to accept or decline the offer; (4) that the consequence of accepting the offer is that both the agent and the participants will be compensated in accordance with the agent's proposition, and (5) that the consequence of rejecting the offer is that neither the agent nor the participant will receive any proportion of the pool of money.

After providing the participant information about the rules of the ultimatum game, the agent proceeded to give an offer which was fixed at 20 SEK and instructed them to press the green button on the input device, placed right before them to accept the offer or the red button to reject the offer. After the interaction session, participants were asked to fill out a questionnaire which consisted of the scale measuring social presence. Finally, participants that had accepted the offer were compensated with the amount of 20 SEK.
Chapter 4

RESULTS

4.1 Effects of embodiment

The effect embodiment on social influence. A chi-square test for association was conducted between agent condition and participants’ decision to accept or reject the offer made by the agent. All expected cell frequencies were greater than five. Participant responses between agent conditions were not statistically different, \( \chi^2(1) = 2.411, p = .121 \). This result indicates that participants’ decision to accept or reject the offer made by the agent was not influenced by the agent’s physical or virtual presence.

The effect of embodiment on prosocial behaviour. An independent-samples t-test was conducted to compare the number of completed math exercises in physical and virtual agent conditions. Five outliers were identified in the data, as assessed by inspection of a boxplot. Outlier data points (12 math exercises or more) were altered and coded as 9, the second largest value found in the data. There was homogeneity of variances, as assessed by Levene’s test for equality of variances \( p = .477 \), however number of math exercises was still not normally distributed after dealing with outliers, as assessed by Shapiro-Wilk’s test \( p < .001 \). Number of completed math exercises was not statistically different between the physical agent condition \( 4.9 \pm 2.55 \) and the virtual agent condition \( 5.63 \pm 3.30 \), \( t(58) = -.685, p = .496 \).

A Fisher’s Exact test was conducted between agent condition and participants’ decisions whether to donate the accepted amount of money back to the agent. Participants’ decisions between agent conditions were not statis-
tically different, \( p = 1.000 \).

Taken together, these results indicate that agent embodiment did not affect participants’ inclination toward prosocial behaviour (in this case helping with math exercises and donating money).

**The effect of embodiment on perceived social presence.** A Mann-Whitney U test was run to determine if there were differences in perceived social presence between participants in physical and virtual agent conditions. Distributions of the social presence scores for the two groups were not similar, as assessed by visual inspection. Median social presence score for participants in the physical agent condition (mean rank = 28.9) and participants in the virtual agent condition (mean rank = 32.1) was not statistically different, \( U = 498, z = .703, p = .482 \). This result indicates that we should retain the hypothesis that there is no difference between physical and virtual embodiment and presence in terms of the social presence elicited by agent with these attributes.

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Social influence</th>
<th>Prosocial behaviour</th>
<th>Social presence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virtual EMB</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Summary.** These results point in favour of rejecting hypotheses EMB1–EMB3 (Table 4.1) which stated that the physical agent would be more socially influential (EMB1), elicit a higher level of prosocial behaviour in participants (EMB2), and perceived of as socially present to a higher the degree than the virtual agent (EMB3). These result clearly conflict with previous research on physical versus virtual agent embodiment, which point toward in general more favourable responses from interacting with physical than virtual agents (see section 2.3). However, the validity of the results regarding the testing of EMB1–2 are brought to doubt due to an failure to use sufficiently large sample sizes for binomial data, which may have resulted in type II errors. This matter can be attributed to deficient experiment design, as will be further discussed in the limitations section of the discussions chapter.
4.2  Effects of perceived social presence

The effect of perceived social presence on social influence. A Mann-Whitney U test was run to determine if there were differences in perceived social presence between participants that accepted and rejected the offer made by the agent. Distributions of the social presence scores for the two groups were similar, as assessed by visual inspection. Median social presence score was statistically significantly higher in participants that accepted the offer (6.69) than in participants that rejected the offer (5.94), $U = 610$, $z = 2.395$, $p = .017$. This result indicates that the perceived level of social presence is positively related to the social influence of the robot on the decision of the participants to accept or reject the agent’s offer. Participants that experienced the agent as a highly social actor were more inclined to accept the offer than participants that did not.

The effect of perceived social presence on prosocial behaviour A Mann-Whitney U test was run to determine if there were differences in perceived social presence between participants that donated the accepted amount of money back to the agent and those who did not donate the money. Distributions of the social presence scores for the two groups were similar, as assessed by visual inspection. Median social presence scores were not statistically significantly different between those two donated money (6.88) and those who did not (6.5), $U = 162.5$, $z = 1.866$, $p = .061$.

Table 4.2

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Social influence</th>
<th>Prosocial behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td>High SP</td>
<td>Sig. dif.*</td>
<td>Not sign. dif.</td>
</tr>
<tr>
<td>Low SP</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Statistically significant group difference at the level of $p < .05$

Summary. Participants that experienced the agent as a highly social actor were more inclined to accept the offer than participants that did not. No effect was found of participants’ perceived social presence on participants’ prosocial behaviour.
4.3 Negative attitude and social presence

The relationship between negative attitude and perceived social presence. A Mantel-Haenszel test of trend was run to determine whether a linear association existed between negative attitude toward robots and perceived social presence. Negative attitude toward robots were scored from 1 to 5 (with higher numbers indicating higher degrees of negativity) and social presence was scored from 1 to 10 (with higher numbers indicating higher degrees of perceived social presence). No statistically significant linear association between negative attitude and social presence was found, $\chi^2(1) = 276.1, p = .573, r = -.074$. This result indicates that the experience of an artificial agent as a social actor is not, as was hypothesised, related to positive attitude toward it.

<table>
<thead>
<tr>
<th>Correlate variables</th>
<th>Negative attitude</th>
<th>Social presence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative attitude</td>
<td>-</td>
<td>No correlation</td>
</tr>
<tr>
<td>Social presence</td>
<td>No correlation</td>
<td>-</td>
</tr>
</tbody>
</table>

Summary. This result points toward the rejection of hypothesis ‘NASP’, which stated that there is a negative correlation between negative attitude and perception of the agents as highly socially present. Instead, these results suggest that the experience of the agents as social actors were not related to having positive or negative attitudes toward them.

4.4 Gender effects

The effect of gender on negative attitude. A Mann-Whitney U test was run to determine if there were differences in negative attitude between male and female participants. Distributions of the general negative attitude scores for the two groups were similar, as assessed by visual inspection. Median general negative attitude score was moderately statistically significantly higher for women (3.21) than for men (2.86), $U = 524, z = 1.949, p = .051$. 

34
The effect of gender on perceived social presence. A Mann-Whitney U test was run to determine if there were differences in perceived social presence between male and female participants. Distributions of the general negative attitude scores for the two groups were similar, as assessed by visual inspection. Median social presence score was statistically significantly higher for women (7.00) than for men (6.00), $U = 569, z = 2.653, p = .008$.

The effect of gender on perceived social influence. A chi-square test for association was conducted between gender and participants’ decision to accept or reject the offer made by the agent. All expected cell frequencies were greater than five. Participant responses between agent conditions were not statistically different, $\chi^2(1) = .837, p = .360$. This result indicates that gender did not influence participants’ decision to accept or reject the offer made by the agent.

The effect of gender on prosocial behaviour. A Fisher’s Exact test was conducted between gender and participants’ decisions whether to donate the accepted amount of money back to the agent. There was a statistically significant association between gender and inclination toward donating, $p = .013$, with 100% of female participants donating the money back to the agent compared to 54% of the male participants. This result suggests that female participants exhibited a higher degree of prosocial behaviour when compared with male participants.

An independent-samples t-test was conducted to compare the number of completed math exercises between gender groups. Five outliers were identified in the data, as assessed by inspection of a boxplot. Outlier data points (12 math exercises or more) were altered and coded as 9, the second largest value found in the data. There was homogeneity of variances, as assessed by Levene’s test for equality of variances ($p = .347$), however number of math exercises was still not normally distributed after dealing with outliers, as assessed by Shapiro-Wilk’s test ($p < .05$). Number of completed math exercises was not statistically different between male participants ($4.83 \pm 2.22$) and female participants ($5.40 \pm 2.86$), $t(58) = -.859, p = .394$.

Summary. Female participants rated the artificial agents as having a higher degree of social presence than men while also rating them more negatively than male participants in the negative attitude scale. Indicating that the experience of an artificial agent as a social actor is not related to having
positive or negative attitudes toward it, this result provides evidence toward the rejection of hypothesis ‘NASP’ which stated that there would be a negative correlation between negative attitude and perceiving of the agent as highly socially present. There was no gender effect with regard to the social influence of the agent on participants’ decision to accept or reject the offer made by the agent. Female participants exhibited a higher degree of prosocial behaviour, with a higher percentage of female participants donating money than male participants.

4.5 Conclusion

None of the effects proposed in hypotheses EMB1–EMB3, regarding robotic versus virtual embodiment, were observed in hypothesis testing. EMB1 and EMB2 stated that there is a larger effect of physical than virtual embodiment on social influence and prosocial behaviour, respectively. The failure to use sufficient sample sizes for binomial data in testing these hypotheses renders causal inferences largely unwarranted, considering the relatively high risk of committing a type II error (see ‘Limitations’ in the discussion chapter). Results point in favour of rejecting EMB3, which stated that there is a larger effect of physical than virtual embodiment on participants’ perceived social presence of the agent.

Hypotheses SP1 and SP2 stated that agents are more influential and elicit higher degrees of prosocial behaviour when perceived as highly socially present than when perceived as socially present to a lesser degree. Evidence was found in favour of SP1 (effect of social presence on social influence) but not for SP2 (effect of social presence on prosocial behaviour). The results also point in toward rejecting hypothesis NASP, which stated that negative attitude and perceived social presence correlate negatively.

Finally, there were differences between female and male participants in their negative rating of the agents and in their tendencies for prosocial behaviour. On average, Females rated the agents more negatively and exhibited a higher degree of prosocial behaviour than men.
Chapter 5

DISCUSSION

Findings regarding the relation between agent embodiment and social presence and influence is presented in Figure 5.1.

Figure 5.1: Statistical results concerning the link between agent embodiment, presence and influence

The first thing to note is that the results of the experiment show that participants’ behaviour was influenced by the extent to which they perceived of the agent as a social actor (i.e. ‘social presence’). This indicates that social presence is a factor that needs to be taken into account in the design of human-agent social interaction, especially in cases where agent influence is considered to be important. Contrary to what was expected, there was no difference in the extent to which the physical and virtual agents were per-
ceived of as socially present or in the influence they exerted on participant behaviour. This suggests that agent embodiment is not strongly related to perceived agent presence or social influence in the context of the type of interaction scenario that was featured in the present experiment. It may be the case that there are larger effects of embodiment in contexts where having a physical (or a virtual) body is crucial for successfully executing certain tasks. However, in our experiment both agents interacted exclusively verbally with participants, and there may have been no particular benefits to physical or virtual embodiment in this context. In other words, the social presence and influence of artificial agents may be defined at large by the nature of the task domain and how well the agent is able to engage with it.

The results from a study on agent embodiment by Hasegawa, Cassell and Araki (2010) lend some credibility to this interpretation. This study evaluated effects on participants’ performances and attitudes of interacting with three different ‘direction-giving systems’: a robot, a virtual agent and a GPS system. All three system provided directions to locations around a campus area verbally. In addition to verbal behaviour the robot and the virtual agent were capable of indicating directions by using gestures, and the GPS system could display a map on a screen. The experiment assessed various effects of embodiment (agent type) and gestures (present or absent). The results showed that participants preferred the embodied agents over the GPS-system, and that they preferred the robot over the virtual agent only when gestures were present in the behaviour of the agents (i.e. there was no preference when gestures were absent). Here agent embodiment was significant to the interaction in particular when an agent-specific quality (in this case gesturing) was employed. Presumably, the physical agent gestures were more easily interpreted than virtual agent gestures because of the two-dimensional limitations of the latter agent. The authors concluded that:

The comparison between visible maps (GPS systems) and human-like behavior in embodied agents is not as straightforward as it has been portrayed, nor is the comparison between robots and [virtual agents] a clear win for either side. What is clear is that the details of embodiment and their relationship to task and talk are central to the design of embodied systems. (ibid., p. 31)

If the above line of reasoning is correct, then conclusions drawn from studies that investigate effects of embodiment, by means of comparing different types of agents, should not be over-generalised beyond the particular
5.1 Limitations

Hypothesis testing showed no effects of physical versus virtual embodiment on the social influence of the agent on participant behaviour. However, these results were based on statistical hypothesis testing of a binomial data sample of 60 participants. Given the assumption that no previous studies have investigated the effect of physical versus virtual embodiment on social influence in an ultimatum game it is reasonable to expect that a proportion of .5 of the participants will accept the offer and that a corresponding proportion of .5 will decline the offer. According to Sauro and Lewis (2012), a reasonable sample-size estimation of binomial data, with an expected proportion of .5, a desired level of power of .05, and a desired level of confidence of .95, is 385. The sample size \( (N = 60) \) used in the present study to measure social influence was thus considerably lower than this recommendation. This study design flaw is likely to have adversely affected the validity of the results from the testing of social influence in EMB1 (response to offer) and EMB2 (donation request). A better approach for measuring these agent qualities would be to use a parametric or ordinal instead of a nominal metric.

The present study assessed various effects of physical versus virtual agent embodiment by comparing a co-present robot with a virtual agent. Such comparisons may conflate what is arguably two distinct dimensions of artificial agent embodiment, namely the attribute of having a physical or virtual body and the attribute of being physically or virtually present (see Li, 2015). The argument is that agents that have a physical body can be either physically present (co-located) or virtually present (tele-present), and that comparisons between co-present robots and virtual agents therefore conflate body type with presence by comparing agents that have a physical body and are physically present (i.e. robots) with agents that have virtual bodies and are virtually present (i.e. virtual agents). See Table 5.1 for an illustration of this characterisation of agent types.
Table 5.1  
*Agent classification found in Li (2015)*

<table>
<thead>
<tr>
<th>Body type</th>
<th>Physical</th>
<th>Virtual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical</td>
<td>Robot</td>
<td>Tele-present robot</td>
</tr>
<tr>
<td>Virtual</td>
<td>-</td>
<td>Virtual Agent</td>
</tr>
</tbody>
</table>

The methodological recommendation made by Li (2015), who advocates this type of agent classification, is that studies that investigate effects of physical or virtual body type should compare tele-present robots with virtual agents (agent presence is fixed in both experimental conditions). Similarly, studies that investigate effects of physical or virtual presence should compare co-present robots with tele-present robots (agent body type is fixed in both experimental conditions).

Whereas there are methodological merits to the above argument, the decision made in this study to compare a co-present robot with a virtual agent is in part motivated by the fact that tele-present agents are used in a considerably narrower set of operational domains than the other two types of agents. Co-present robots, tele-present robots, and virtual agents, all bring to the table different qualities and opportunities for successful social interaction. However, from a design standpoint the choice is most often between a co-located robot or a virtual agent, probably since the use of tele-present robots often can be replaced with virtual agents with the benefit of cutting expenses related to robots. Co-present robots, in contrast, are often used for various physical tasks and therefore cannot always be replaced by virtual agents. The decision to compare a co-present robot with a virtual agent is thus motivated in part from a pragmatic viewpoint.

Another rationale for this decision is that the assumed theoretical basis for the above agent classification is questionable. Specifically, the notion of physical versus virtual presence and the classification of virtual agents as virtually present is dubious. Physical presence is defined here as ‘being in the same physical space as the user’ and virtual presence is defined as ‘being shown as a live video feed on a screen or projection’ (Li, 2015, p. 25). It makes sense to predicate physical presence to co-present robots, and virtual presence to tele-present robots given these definitions. However, it is unclear what it means for virtual agents to be virtually present. Virtual agents are neither present in the sense of appearing to be close but actually being far away which is implicit in the above definition of virtual presence.
as being shown as something (which one is not) nor share the same physical space as the user.

The present study is also affected by general limitations of studying social interaction with artificial agents in laboratory experiment settings. One such limitation is the increased risk of experimenter effects biasing the results. Such effects are hard to measure or minimise, since they are hard to predict and can sometimes be unavoidable. It is reasonable to assume that the experimenter influenced to some extent in the context of this study. There are however no obvious reasons to expect that the experimenter exerted different levels of influence across the group comparisons in the experiment, since the experimenter carried out the experimental procedure in the same manner in all trials. Furthermore, the artificiality of laboratory settings may produce unnatural user responses that do not reflect real life. The ecological validity of evaluative studies of artificial agents is important since the ‘true benchmark’ of artificial agents is how well the agent performs in the task domain that it was built for, and the task domains are potentially very diverse. A risk of ignoring evaluations of artificial agents ‘out in the wild’ is that the evaluation ‘misses its target’ either by assessing irrelevant features of the agent or failing to assess important features of the agent. One way of mitigating this effect, which is suggested in the section of this thesis called ‘The making of social agents’, is to establish conventions to write up formal descriptions of the agents’ intended task domain, social qualities, morphologies, and so on, that can then function as a roadmap in evaluation.

It is duly said here that while there are drawbacks to studying social interaction with artificial agents in the laboratory, there are also tremendous advantages due to the benefits of experimental control. While validating social sufficiency in realistic ecologies could prove beneficial to the development of artificial agents, it is very hard to infer the causes of social sufficiency without controlled experimental approaches where effects of particular agent attributes on the social capacity of the agent can be assessed systematically.

### 5.2 Future studies

Although there are evidence in favour of robotic embodiment being more beneficial to social interaction overall, comparisons between robots and virtual agents are not a clear win for either side. Studies compare different agents are by large only informative of the particular type of interactions investigated. This study concluded that there were no effects of embodiment
on agent social influence or presence in the context of an ‘ultimatum game’ decision making task. This represents a valuable but small contribution in the research domain of human-machine social interaction. Social reality is rich, varied and complex. Creating artificial social agents that fit into it is not easy, and we have probably just begun to explore the place for them. It is reasonable to expect that there are – as in the case of different animal bodies – advantages and disadvantages to robotic and virtual embodiments in social interaction, and that these depend on the nature of the situation of interaction.

The finding that the social influence of the agents was influenced by the extent to which participants perceived of the agents as social actors may be a relationship that transcends particular contexts of interaction. In other words, social influence may be an inexorable effect of social presence. This idea resonates with the observation that our perceptions of agents as socially present, and the social influence that they exert on us, seem to strongly correlate in nature. Agents that are perceived of as non- or semi-social but not highly social, such as most animals and robots, exert lesser degrees of social influence on us. Indeed, it is hard to imagine, *ceteris paribus*, an entity with rich social faculties that possesses no social influence on us whatsoever. It is clear that the notion of sociality, and the relationship between social presence and influence, is important to the study and development of robots and virtual agents.
References


Heider, F., & Simmel, M. (1944). An Experimental Study of Apparent Be-


