Responses to human-like artificial agents : effects of user and agent characteristics

Responses to Human-like Artificial Agents

Peter A. M. Ruijten

The work in this dissertation was part of a long-term energy research (EOS) project with the title ‘Persuading households to save energy through smart agents’, funded by the Netherlands Enterprise Agency.

A catalogue record is available from the Eindhoven University of Technology Library.

ISBN: 978-94-6295-167-9
NUR: 775
Keywords: Anthropomorphism || Persuasive Technology || Artificial agents

Cover design: Peter Ruijten & Harrie de Weert
Printed by: Proefschriftmaken.nl || Uitgeverij BOXPress.
Published by: Uitgeverij BOXPress, ‘s Hertogenbosch

Responses to Human-like Artificial Agents; effects of user and agent characteristics

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus prof.dr.ir. F.P.T. Baaijens, voor een commissie aangewezen door het College voor Promoties, in het openbaar te verdedigen op donderdag 21 mei 2015 om 16.00 uur

door

Petrus Adrianus Maria Ruijten

geboren te Oud en Nieuw Gastel

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter:	prof.dr. I.E.J. Heynderickx
1e promotor:	prof.dr. C.J.H. Midden
co-promotor:	dr.ir. J.R.C. Ham
leden:	jprof.dr. F. Eyssel (Universität Bielefeld)
	prof.dr. E.O. Postma (Universiteit Tilburg)
	prof.dr. E.M. Steg (Rijksuniversteit Groningen)
	prof.dr. P. Markopoulos

Table of contents

1 General introduction **5**

1.1 Energy conservation 7

1.2 Persuasive Technology 8

1.2.1 Advantages of Persuasive Technology 8

1.2.2 Persuasive Technology in the social actor role . . 9

1.2.3 Social cues in Persuasive Technology 10

1.3 Anthropomorphism 11

1.3.1 Interpretations of anthropomorphism 11

1.3.2 Elements of anthropomorphism 13

1.4 Scope of this dissertation 17

2 Lonely and susceptible **21**

2.1 General introduction 22

2.1.1 Human-likeness of non-human agents 23

2.1.2 The need to belong 24

2.1.3 Research aims 25

2.2 Study 1 26

2.2.1 Method 26

2.2.2 Results 31

2.2.3 Discussion 35

2.3 Study 2 38

2.3.1 Method 38

2.3.2 Results . . . 41
2.3.3 Discussion . . . 47
2.4 General discussion . . . 49
2.4.1 Limitations and future research . . . 52
2.4.2 Conclusions . . . 53

3 Ambiguous agents **55**
3.1 General introduction . . . 56
3.1.1 Types of social cues . . . 57
3.1.2 Consistency of social cues . . . 58
3.1.3 Research aims . . . 59
3.2 Study 3 . . . 60
3.2.1 Method . . . 62
3.2.2 Results . . . 64
3.2.3 Discussion . . . 68
3.3 Study 4 . . . 70
3.3.1 Method . . . 71
3.3.2 Results . . . 73
3.3.3 Discussion . . . 76
3.4 General discussion . . . 78
3.4.1 Limitations and future research . . . 79
3.4.2 Conclusions . . . 80

4 Measuring Perceived Human-likeness **83**
4.1 General introduction . . . 84
4.1.1 Subsets of human-like characteristics . . . 85
4.1.2 Two-dimensional approaches . . . 86
4.1.3 Conceptualizing anthropomorphism . . . 88
4.1.4 The Rasch model . . . 90
4.1.5 Research aims . . . 91
4.2 Study 5 . . . 92
4.2.1 Method . . . 93

4.2.2 Results and discussion 95
4.2.3 Conclusions 102
4.3 Study 6 102
4.3.1 Method 103
4.3.2 Results and discussion 105
4.3.3 Conclusions 111
4.4 Study 7 112
4.4.1 Method 112
4.4.2 Results and discussion 114
4.4.3 Conclusions 122
4.5 General discussion 123
4.5.1 Unidimensionality of anthropomorphism 124
4.5.2 Differences between the characteristics 124
4.5.3 Comparison of measuring instruments 125
4.5.4 Limitations and future research 126
4.5.5 Conclusions 127

5 General discussion 129
5.1 The role of anthropomorphism 131
5.1.1 A different mediating process 132
5.1.2 Effects too small to detect 133
5.1.3 Issues with available measuring instruments . . . 134
5.2 Predispositions to anthropomorphize 135
5.3 Contributions of the current work 136
5.4 Recommendations of the current work 138
5.5 Limitations and future work 139
5.6 Ethical considerations 143
5.7 Conclusions 144

References 145

Summary 165

Samenvatting 167

List of Figures 169

List of Tables 173

Appendices 175

A Questionnaires 177

A.1 Godspeed-instrument 178
A.2 Waytz-instrument 179
A.3 Game-evaluation 180
A.4 Perceived intelligence 181
A.5 Perceived agent knowledge 182
A.6 Need to belong 183
A.7 Agent-evaluation 184
A.8 Predisposition to anthropomorphize 185

B Washing tasks 187

C Heating tasks 191

Acknowledgments 195

Curriculum Vitae 199

Publications 201

Colophon 203

"There is a universal tendency among mankind to conceive all beings like themselves, and to transfer to every object, those qualities, with which they are familiarly acquainted, and of which they are intimately conscious."

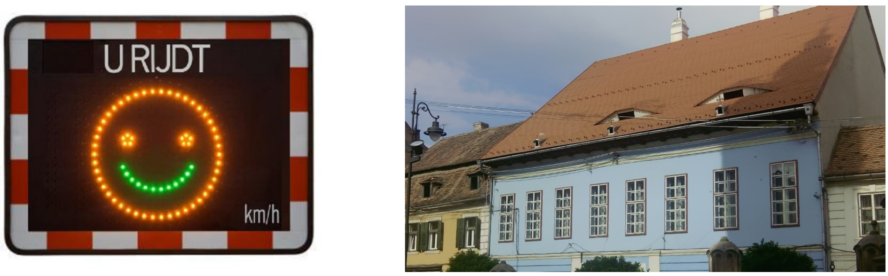
David Hume (1711 – 1776)

1 General introduction

Technology has become a major part of our daily lives, as it has entered almost every arena of human society. Computers operate everywhere from people's homes and the streets to their workplaces. They come in so many shapes and sizes that it is sometimes difficult to recognize them. They have become more intrusive, more intimate, and a bigger part of our social lives, which has made them more like assistants or buddies than objects. Earlier work in the domain of human-computer interaction showed that people interact with computers in similar ways as they interact with other humans (Reeves & Nass, 1996). This effect may be caused by people perceiving computers as similar to other humans.

When computers are perceived as similar to other humans, they might also influence their users' behavior in similar ways in which humans influence each others' behavior. Technology that is designed to influence its users' behavior is called Persuasive Technology. This type of technology could benefit from being perceived as similar to humans. This may be why some road signs that prevent drivers from exceeding the speed limit show faces rather than numbers (see Figure 1.1a).

Humans have a general tendency to perceive human-like characteristics in (i.e. anthropomorphize) non-human objects. For example, humans seem to be hardwired for detecting faces (e.g., Perrett et al., 1983). We see them in buildings that have windows that look like eyes (see Figure 1.1b), or even in clouds (for examples of how this perceptual strategy characterizes human experiences, see Guthrie, 1993).



(a) (b)

Figure 1.1: Examples of objects that resemble human-like appearances, with (a) a road sign that provides information about a vehicle's speed with a smiley face and (b) a house that appears to have eyes.

Could perceived human-likeness increase a Persuasive Technology's persuasiveness? The current dissertation aims to answer this question by examining the relation between perceived human-likeness of Persuasive Technology and its ability to influence its user's behavior.

The work in this dissertation was part of a project for promoting energy conservation, and it was aimed at designing persuasive interventions to stimulate sustainable behavior. In the next section, I will describe this energy conservation context.

1.1 Energy conservation

Energy conservation can be approached in three different ways, with the focus being on either technological developments, human behavior, or interactions between people and technology. Studies have shown that purely focusing on *technological* factors of energy conservation leads to disappointing results, mainly because of so called rebound effects (e.g., Khazzoom, 1980; Hertwich, 2005; Midden, Kaiser, & McCalley, 2007; Herring & Roy, 2007). More specifically, improvements in energy efficiency could make people use their technological products more often, leading to *higher* consumption levels (Herring & Roy, 2007). In other words, strategies aimed at reducing energy consumption by developing more energy-efficient products might have the opposite effect. It is therefore important to acknowledge the role of human behavior in energy consumption.

Human behavior is argued to be one of the most important determinants of energy conservation, because it is this behavior that largely determines a person's energy consumption (e.g., Stern, 1992; Lutzenhiser, 1993; Jaffe & Stavins, 1994; Abrahamse, Steg, Vlek, & Rothengatter, 2007; Webler & Tuler, 2010). This behavior could be influenced by *psychological* factors of energy conservation which are aimed at changing a person's perception and cognition related to energy use. This is what Steg (2008) referred to as psychological strategies in her framework on energy-efficiency. Psychological strategies particularly focus on individuals' psychological processes that make them more likely to choose for sustainable options than unsustainable ones.

Rather than focusing on either technological or psychological factors, energy consumption could also be viewed as a result of *interactions* between people and the technology they use (see e.g., Midden et al., 2007; Midden, McCalley, Ham, & Zaalberg, 2008). Most, if not all, energy-related behavior occurs in interactions between people and their technological systems, so it is in these interactions where the crucial behavioral decisions are made. Persuasive interventions aimed at reducing energy use could thus be most effective within those interactions. For this reason, user interfaces of energy-related appliances could be designed with the purpose of influencing people's energy conservation behavior. Such technological systems are referred to as Persuasive Technology.

1.2 Persuasive Technology

The study of Persuasive Technology began in the 1990s and it has grown into a global area of research and design within the human-computer interaction domain. Persuasive Technology (PT) is defined as *"a computing system, device, or application intentionally designed to change a person's attitudes or behavior in a predetermined way"* (Fogg, 1999, pp 27). It should be designed to facilitate behavior changes without taking away control from its users (Fogg, 1999, 2003). Research has shown that PT effectively influences people's attitudes and their sustainable behavior (see e.g., Arroyo, Bonanni, & Selker, 2005; Midden et al., 2008; Lockton, Harrison, & Stanton, 2008; Zapico, Turpeinen, & Brandt, 2009; Kappel & Grechenig, 2009; Petkov, Köbler, Foth, & Krcmar, 2011).

1.2.1 Advantages of Persuasive Technology

PT has several advantages compared to human persuaders. For example, computers are more persistent than humans, they have the option of remaining anonymous, have the capability of storing huge amounts of data, and they are ubiquitous in the sense that they can be (almost) anywhere, including private areas (for an overview, see Fogg, 2003). The benefit of being used everywhere creates great opportunities for applications in the context of energy conservation. In particular, technology that provides immediate visible and comprehensive information about the consequences of a person's behavior on his/her energy use could effectively reduce that energy use (Wood & Newborough, 2003). For example, adding information to a thermostat interface about the consequences of increasing the temperature settings could make a person use less energy.

PT can be categorized into three functional roles. As a tool, it can increase people's ability to perform a certain behavior, for example by ambient displays that represent a person's computer use or his/her physical activity (see e.g., T. Kim, Hong, & Magerko, 2010; Burns, Lueg, & Berkovsky, 2013). As a medium, PT can provide experiences that express a point of view, for example by simulating flooding experiences for increasing people's awareness of climate change (see e.g., Zaalberg & Midden, 2010). As a social actor, PT can even engage in social interactions, create relationships, and evoke social responses (Fogg, 2003). PT in the role of a social actor could be very powerful in influencing people's behavior. The work in this dissertation will utilize PT as a social actor for influencing people's energy-related behavior.

1.2.2 Persuasive Technology in the social actor role

In the social actor role, PT can be especially effective when it has human-like physical features or emotions (Fogg, 1999). Examples of PT that includes human-like features are artificial agents and robots (e.g., Sproull, Subramani, Kiesler, Walker, & Waters, 1996; Bailenson & Yee, 2005; Kidd & Breazeal, 2004; C. Kim & Baylor, 2008; Midden & Ham, 2009; Siegel, Breazeal, & Norton, 2009; Midden & Ham, 2012). Such artificial agents or robots could influence people's behavior through social interactions. For example, social feedback provided by a robotic agent showing emotional expressions and saying things like 'well done' made people save more energy than factual feedback provided by an energy bar (Midden & Ham, 2009). This effect was in accordance with earlier findings that showed that adding social approval (with a smiling face) or disapproval (with a frowning face) to information about energy consumption made people save more energy than only providing (factual) information (Schultz, Nolan, Cialdini, Goldstein, & Griskevicius, 2007). Based on these findings, I argue that by successfully adopting the role of a social actor, PT could increase its effectiveness in influencing people's attitudes and behavior. One approach to increase the socialness of PT is by including more social cues in its design.

1.2.3 Social cues in Persuasive Technology

During the last 20 years, various studies have demonstrated that adding social cues to PT could increase its persuasiveness (e.g., Nass, Steuer, Tauber, & Reeder, 1993; Nass, Steuer, & Tauber, 1994; Fogg, 2003; Nass & Brave, 2005; Midden et al., 2008; Vossen, Ham, & Midden, 2009; Midden & Ham, 2013). In spite of these positive results, the relation between perceived human-likeness and showing social cues does not seem to be linear. That is, more social cues do not necessarily lead to more perceived human-likeness. Instead, they create opportunities for applying social influence strategies, because people respond to technology that shows social cues in similar ways as they respond to other people (Reeves & Nass, 1996). For example, when a technology uses human-like language or flattery, it evokes social responses (for an overview, see Reeves & Nass, 1996).

Only few studies have investigated how people perceive combinations of social cues and whether those combinations strengthen or weaken each other. As I will show in this dissertation, multiple social cues could also create pitfalls when those cues do not match, because this could change the combined meaning of those cues making them being perceived as less rather than more human-like.

In sum, adding social cues to the design of PT may influence its perceived human-likeness, and evoke automatic (i.e. unconscious) social responses. These automatic responses can be classified as results of attributions of human-like characteristics to that technology (i.e. anthropomorphism).

1.3 Anthropomorphism

When David Hume presented his view on people's tendency to conceive all beings like themselves in the book The Natural History of Religion (Hume, 1889), he was one of the first to describe the concept anthropomorphism. Anthropomorphism is generally defined as the tendency to attribute human-likeness to non-humans (e.g., Guthrie, 1993; Nass & Moon, 2000; Fong, Nourbakhsh, & Dautenhahn, 2003; Duffy, 2003; Epley, Waytz, & Cacioppo, 2007; Bartneck, Kulić, Croft, & Zoghbi, 2009). This definition is quite abstract, and different researchers appear to have different interpretations of the concept.

1.3.1 Interpretations of anthropomorphism

When Duffy (2003) defined anthropomorphism as *"attributing cognitive or emotional states to something based on observation in order to rationalise an entity's behaviour in a given social environment"* (pp 180), he emphasized that anthropomorphism is a result of observing the behavior of non-humans. This view is in line with that of Bartneck et al. (2009) who described anthropomorphism as a result of realistic human-like appearances. Both Duffy (2003) and Bartneck et al. (2009) thus focused on observations of *technological* features that could trigger anthropomorphic responses.

Other researchers have focused on *psychological* states and processes that could trigger anthropomorphic responses (e.g., Guthrie, 1993; Epley et al., 2007). These authors argued that anthropomorphism is an automatic unconscious process, and consequently, they focused on how a user's psychological states and processes could trigger anthropomorphic responses (for an overview, see Epley et al., 2007). These psychological states and processes have been studied extensively in various contexts, and several studies have investigated the extent to which individual differences in personality could make people attribute human-like characteristics to technology or other objects (e.g., Epley et al., 2007; Epley, Waytz, Akalis, & Cacioppo, 2008; Waytz, Cacioppo, & Epley, 2010; Waytz, Epley, & Cacioppo, 2010; Waytz, Morewedge, et al., 2010; Waytz, Heafner, & Epley, 2014).

The argument that anthropomorphism is an unconscious process is shared by Nass and Moon (2000). They however referred to this concept as 'ethopoeia', which they defined as *"a direct response to an entity as human while knowing that the entity does not warrant human treatment or attribution"* (Nass & Moon, 2000, pp 94). In fact, they referred to the term anthropomorphism as involving thoughtful, sincere beliefs that a technology has human-like characteristics. So, even though there appears to be some disagreement about which term best describes people's social responses to technology, Nass and Moon (2000) agreed with aforementioned authors that the process itself is an unconscious one.

In addition to focusing on technological or psychological determinants, anthropomorphism has also been described as a mechanism that controls people's responses to or interactions with something that appears to contain human-like elements. As Fong et al. (2003) argued, *"the role of anthropomorphism is to function as a mechanism through which social interaction can be facilitated"* (pp 150). With social interactions they referred to interactions between people and technology (or other objects) that are more or less perceived as human-like (Fong et al., 2003).

In sum, there seems to be consensus among researchers that anthropomorphism (or at least the process of attributing human-likeness to technology) is an unconscious process, although the empirical evidence for this is still limited. The work by Roubroeks, Ham, and Midden (2011) demonstrated that people responded socially to computer agents (and objects) at a spontaneous level, but not at an intentional level. These findings suggest that anthropomorphism is an unconscious rather than a conscious process.

Even though anthropomorphism appears to be a process of which people are not aware, they may be consciously experiencing *outcomes* of that process. For example, people might not be aware of the fact that they attribute human-likeness to a computer that greets them, but they can experience changes in their mood or behavior as a result of these attributions. In my view, different outcomes of the process, as well as several other important elements, should be included in the description of anthropomorphism.

1.3.2 Elements of anthropomorphism

I describe anthropomorphism as an automatic cognitive process of attributing human nature or human uniqueness characteristics to nonhuman technological, spiritual or natural objects and entities, causing a mixture of affective, behavioral and cognitive responses. This description has four different elements.

First, I argue that anthropomorphism is automatic, requiring minimal effort and occurring without conscious attention (for an overview of automatic processes, see e.g., Hasher & Zacks, 1979; Bargh, 1984). In other words, people can effortlessly attribute human-likeness to objects, for example when those objects seem to resemble human-like appearances. Important to note here is the difference between automaticity and unconsciousness. These two concepts stem from different lines of research (for an overview, see Payne & Gawronski, 2010). The roots of the concept of automaticity can be found in attention research, where a distinction is made between controlled and automatic modes. Automaticity thus refers to processes that are outside of a person's *control*. The roots of the concept of unconsciousness can be found in implicit memory research, which focuses on a dichotomy between explicit and implicit processes. Unconsciousness thus refers to processes that are outside of a person's *awareness* (Payne & Gawronski, 2010). I argue that, although the process of anthropomorphism is both automatic and unconscious, its *outcome* does not necessarily need to be unconscious. For example, people may experience changes in their mood and behavior as a result of anthropomorphizing a non-human object.

Second, all human-like characteristics that can be attributed to non-humans are human nature or human uniqueness characteristics. Although these types of characteristics are regarded as two distinct understandings of humanness (Haslam, Bain, Douge, Lee, & Bastian, 2005), they both represent characteristics that belong to the notion of humanness. Also, human nature and human uniqueness traits together encompass all properties of humanness (Haslam, Loughnan, Kashima, & Bain, 2008).

Third, the nature of the object which is anthropomorphized can be anything ranging from technological, spiritual, natural, to any other type of entity. The target of anthropomorphism thus refers to anything that is not a human being, including all objects, technologies, animals, and Gods.

Finally, the fourth element describes the outcomes of anthropomorphism, which can be anything ranging from moods to behavior changes, indicating that there is no single type of response that is the same for all effects of anthropomorphism.

I argue that all people have a tendency to attribute human-likeness to non-humans (i.e. their predisposition), and some people are generally more likely to anthropomorphize objects than others. A person's tendency could temporarily be influenced by a psychological state, and different aspects of an object may also cause it to be perceived more or less human-like. Anthropomorphism could thus be triggered by factors related to individuals (i.e. psychological determinants) and by factors related to technology (i.e. technological determinants).

Psychological determinants of anthropomorphism

With *psychological* determinants of anthropomorphism I refer to personality traits or psychological states that make an individual more likely to attribute human-like characteristics to technology (or other non-human objects). Some examples of personality traits that influence a person's tendency to anthropomorphize non-humans are provided by Epley et al. (2007). For example, people with a low need for cognition (i.e. those who are not inclined towards effortful cognitive activities, see Cacioppo & Petty, 1982) were found to be more likely to attribute human-like characteristics to technology (and other non-human objects) than people with a high need for cognition (Epley et al., 2007).

An individual's personality or psychological state could thus determine the probability that he/she attributes human-like characteristics to technology. Could such a psychological state also influence susceptibility to social influence by this anthropomorphized technology? This question will be addressed in Chapter 2, in which I investigated the relation between social exclusion as a psychological determinant of anthropomorphism and a technology's persuasiveness.

Technological determinants of anthropomorphism

With *technological* determinants of anthropomorphism I refer to factors in the design of technology that could increase its perceived human-likeness. Some examples of such design elements are eyes or other facial features that enable the expression of emotions, a voice that is used for communication, and human-like shapes such as arms and legs. These elements can be anything between very subtle, such as buttons that can be perceived as eyes (which make me attribute human-likeness to my own washing machine at home, see Figure 1.2) to very unsubtle, such as an interactive artificial agent that shows various emotional expressions and uses gestures and speech. The more human-like elements are included in the design of technology, the more they provide opportunities for it to be anthropomorphized.

Figure 1.2: The washing machine in my home which, by the design of the buttons, makes me attribute human-likeness to it.

A few simple human-like elements in the design of technology could thus make a person attribute a variety of human-like characteristics to it. Could a certain *combination* of human-like elements also influence a person's susceptibility to social influence by that anthropomorphized technology? This question will be addressed in Chapter 3, in which I investigated the relation between consistency of social cues as a technological determinant of anthropomorphism and a technology's persuasiveness.

Combining both determinants of anthropomorphism

When Fogg (2003) discussed effects of PT in the role of a social actor, he merely focused on technological determinants, largely neglecting users' psychological traits that may control their responses to that technology. On the other hand, Epley and colleagues investigated how psychological traits caused people to attribute human-like characteristics to nonhuman objects, but they only focused on characteristics of the user (e.g., Epley et al., 2007; Epley, Waytz, et al., 2008).

It seems to be evident that characteristics of both technology and its users can trigger anthropomorphic responses. These anthropomorphic responses could influence the persuasiveness of that technology. I argue that *anthropomorphism of* and *persuasion by* PT in the role of a social actor are related, and this relation will be investigated throughout this dissertation.

1.4 Scope of this dissertation

The main research question in this dissertation was: What is the relation between anthropomorphism of artificial agents and their persuasive effects? The work in this dissertation was performed as part of a long-term research project titled ‘Persuading households to save energy through smart agents’. For this reason, persuasive effects of artificial agents were investigated in the context of energy consumption. In particular, I examined how technological and psychological determinants of anthropomorphism could influence the persuasiveness of artificial agents that are designed to make their users save energy. In addition, I explored issues regarding the measurement of anthropomorphism and developed a tool for reliable measurement of the concept.

In Chapter 2, I investigated the role of social connectedness as a *psychological* determinant of anthropomorphism (e.g., Epley et al., 2007). When people feel socially isolated, they have a stronger tendency to attribute human-like characteristics to technology, which in turn increases the persuasiveness of that technology. Two studies were performed to investigate effects of being socially excluded on people’s attributions of human-like characteristics to an artificial agent and their susceptibility to persuasion by that agent. In these studies, participants performed an energy-saving task while receiving social feedback from an artificial agent. Participants were either socially included or excluded by playing a short computerized ball-tossing game (i.e. Cyberball, see Williams & Jarvis, 2006). Results of the studies in this chapter highlight the importance of including a user's psychological state in the design of PT in the form of artificial agents.

In Chapter 3, I investigated the role of social cues as *technological* determinants of anthropomorphism (e.g., Nass et al., 1994). Including multiple social cues in the design of PT in the social actor role increases the probability that people attribute human-like characteristics to that technology, which in turn increases its persuasive potential. Two studies were performed to investigate effects of consistency of social cues on people's recognition of an artificial agent's emotions, their attributions of human-like characteristics to the agent, and the agent's persuasiveness. In the first study, participants categorized artificial agents' emotional expressions, and in the second study participants performed an energy-saving task while receiving social feedback from an artificial agent. Results of the studies in this chapter highlight the importance of the combined meaning of social cues for the design of PT in the form of artificial agents.

In Chapter 4, I argue that existing measuring instruments of anthropomorphism may not be sufficient, largely because the concept itself is interpreted differently by different researchers. I propose that human-like characteristics can be mapped onto a one-dimensional scale and ordered according to the probability with which they are ascribed to technology and other objects. A new measuring instrument was developed and tested using the Rasch model, which describes empirical data on a single continuum of equal additive units (see e.g., Bond & Fox, 2013). Results of the studies in this chapter highlight the importance of having a reliable instrument for measuring anthropomorphism.

Finally, in Chapter 5, I discuss the role of anthropomorphism and its relation with PT in more detail, elaborate on the contributions of the presented work, provide recommendations for the design of PT in the social actor role, discuss limitations of the current work, ethical considerations, and draw final conclusions from the findings in this dissertation.

“I would rather take my chance out there on the ocean than to stay here and die on this shithole island, spending the rest of my life talking... TO A GODDAMN VOLLEYBALL!”

Chuck Noland (*Cast Away* – 2000)

2

Lonely and susceptible

The influence of social exclusion and gender on persuasion by an artificial agent

---

This chapter is largely based on:

Ruijten, P. A. M., Midden, C. J. H, and Ham, J. (2014). *Lonely and susceptible: The influence of social exclusion and gender on persuasion by an artificial agent.* Manuscript submitted for publication.

Ruijten, P. A. M., Ham, J., & Midden, C. J. H. (2014). Investigating the influence of social exclusion on persuasion by a virtual agent. In L. Gamberini, A. Spagnolli, & L. Chittaro (Eds.): *PERSUASIVE 2014, LNCS 8462* (pp. 191-200). Springer International Publishing Switzerland 2014.

This chapter explores the relation between a *psychological* determinant of anthropomorphism, social connectedness, and persuasion by an artificial agent. In their theoretical overview, Epley et al. (2007) describe various psychological determinants of anthropomorphism. These sources of anthropomorphism are need for cognition (see Cacioppo & Petty, 1982), need for closure (see Webster & Kruglanski, 1994), desire for control (see Burger, 1992), and chronic loneliness (see Baumeister & Leary, 1995). Chronic loneliness is related to a basic human need, the need to belong, which implies that feelings of social exclusion cause strong effects on emotional patterns and cognitive processes (Baumeister & Leary, 1995). More specifically, people who are lonely feel more anxious and socially isolated than people who are not lonely (Baumeister & Leary, 1995). Moreover, people who are induced to feel socially excluded were more susceptible to a request to donate money than those who were not induced to feel socially excluded (Carter-Sowell, Chen, & Williams, 2008). Thus, social exclusion could lead to increased susceptibility to social influence and thereby increase the persuasive effectiveness of persuasive technology (PT) in the social actor role. In the current chapter, I investigate the relation between social exclusion and persuasion by an artificial agent.

2.1 General introduction

When Chuck Noland was facing physical and social starvation after being stranded alone on a desert island in the movie “Cast Away” (Zemeckis, 2000), he devised strategies to survive both types of challenges. To survive physical starvation, he taught himself how to spear fish, catch rainwater, and find shelter. To survive social starvation, he found companionship in a volleyball he named Wilson. Although not many of us will ever be stranded alone on a desert island, we do share the need for social connection. When Chuck Noland lacked this social connection, he tended to attribute human-like characteristics to a non-human: Wilson the volleyball. Chuck started talking to the volleyball, and he imagined Wilson's responses in their conversations. He actually adapted his behavior to these imagined responses when he decided to first test a gallows he made. This behavior appeared to be a result of humanizing Wilson, which in turn was a result of being socially excluded. This research explores the relation between social exclusion and people's susceptibility to social influence coming from non-human agents. I argue that socially excluded people are more likely to attribute human-like characteristics to such non-human agents than socially included people, and that those who are socially excluded are more susceptible to social feedback from non-human agents than those who are socially included.

2.1.1 Human-likeness of non-human agents

People increasingly interact with technology. These interactions are argued to be similar to interactions with other humans, especially when such technology provides social cues (e.g., Nass et al., 1993, 1994; Reeves & Nass, 1996; Nass & Moon, 2000; Fogg, 2003). Increasing the socialness of such interactions could even induce behavior change. For example, people used less energy in an energy-related task after they were given social feedback by a human-like robot than when they were given factual feedback by an energy indicator (Midden & Ham, 2009). This effect could be explained by the notion that the social feedback increased the robot's perceived human-likeness.

Designing robots and artificial agents in ways that increase their perceived human-likeness could contribute to the naturalness of social interactions with technology (for an overview, see Dalibard, Magnenat-Talmann, Thalmann, et al., 2012). Furthermore, the human-like appearance of robots and artificial agents could enhance their persuasiveness (for an overview, see Baylor, 2009). For example, people showed stronger behavioral intentions to participate in energy conservation as a result of a persuasive message delivered by an anthropomorphic stimulus than by a non-anthropomorphic stimulus (Ahn, Kim, & Aggarwal, 2014). Moreover, anthropomorphic artificial agents have been evaluated more positively (Qiu & Benbasat, 2009), and were experienced as more socially present (Choi, Miracle, & Biocca, 2001) than their non-anthropomorphic counterparts. People also attributed more responsibility to an anthropomorphic interface than to a traditional interface (Johnson, Veltri, & Hornik, 2008).

In sum, anthropomorphic interfaces appear to evoke more social responses than non-anthropomorphic ones, and people generally enjoy the interactions they have with anthropomorphic interfaces more than those with non-anthropomorphic interfaces.

2.1.2 The need to belong

It has been argued that one of the key factors that cause people to attribute human-like characteristics to (i.e. anthropomorphize) non-human objects or technology is the need to belong (Epley et al., 2007). More specifically, the tendency to anthropomorphize objects or artificial agents was shown to be stronger for persons who were chronically lonely compared to those who were chronically connected with others (e.g., Epley, Akalis, Waytz, & Cacioppo, 2008; Waytz, Cacioppo, & Epley, 2010). Furthermore, people who had to recall an event in which they felt lonely tended to anthropomorphize a robot more than people who had to recall random events of the previous day (Eyssel & Reich, 2013). People who were experimentally induced to feel lonely also attributed more humanness to an artificial agent compared to those who were not experimentally induced to feel lonely (Jung & Lee, 2004). Thus, feelings of social exclusion appear to influence anthropomorphic perceptions of artificial agents.

This so called ‘social connectedness’ is argued to be a basic human need (Baumeister & Leary, 1995). When we lack social contact with other people, we feel isolated and adapt our behavior (Baumeister & Leary, 1995). In particular, people who listened to a recorded description that stated that they would end up alone in life or that other participants had rejected them showed impaired performance on cognitive processes (Baumeister, Twenge, & Nuss, 2002), increased aggressive behavior (Twenge, Baumeister, Tice, & Stucke, 2001) and decreased prosocial behavior (Twenge, Baumeister, DeWall, Ciarocco, & Bartels, 2007). Also, people who had to write an essay about a life event in which they felt socially excluded reported higher affiliations with religion (Aydin, Fischer, & Frey, 2010). Furthermore, people high in need to belong were more attentive to social cues than those low in need to belong (Pickett, Gardner, & Knowles, 2004). More specifically, need to belong was found to correlate with people’s accuracy in identifying vocal tone and facial emotions (Pickett et al., 2004).

Together, these findings indicate that socially excluded people are more attentive to social cues and ascribe more human-likeness to artificial agents than socially included people. Such attributions of human-likeness to artificial agents influenced people’s responses to those agents. The question remains whether social exclusion also directly influences people’s susceptibility to social influence.

2.1.3 Research aims

The current research was designed with the aim to investigate the relation between social exclusion and susceptibility to social influence coming from an artificial agent. We hypothesized that a persuasive message from an artificial agent would more strongly influence the behavior of socially excluded people than that of socially included people. This hypothesis was investigated in two studies. Study 1 was designed to investigate whether social exclusion influences susceptibility to persuasion by an artificial agent. Study 2 was designed to replicate the findings of the first study and to investigate to what extent the gender of the artificial agent influences people's susceptibility to persuasion by that agent.

2.2 Study 1

In this study, the effects of social exclusion on susceptibility to social influence coming from an artificial agent were investigated. We hypothesized that social exclusion would increase anthropomorphism of and susceptibility to persuasion by an artificial agent. In addition, a control condition was included to be able to investigate whether social exclusion increases susceptibility or social inclusion decreases susceptibility to social influence.

Earlier work showed that males and females interpret and respond to social exclusion differently (Williams & Sommer, 1997). More specifically, females who were socially excluded compensated by working harder in a collective task, whereas males did not show this tendency (Williams & Sommer, 1997). Therefore, effects of participant gender were also explored in the current study.

2.2.1 Method

Participants and design

Sixty-three participants (38 males and 25 females; $M_{age} = 26.1$, $SD_{age} = 11.1$, Range = 18 to 59) were randomly assigned to one of three experimental conditions of a between-subjects design. Two of the three conditions were designed to make participants feel either socially included ($n = 21$, 14 males and 7 females) or socially excluded ($n = 22$, 14 males and 8 females). The third condition was the control condition, in which no social exclusion manipulation took place (n = 20, 10 males and 10 females). The experiment lasted 30 minutes for which participants were paid €5.

Materials and procedure

At the start of the experiment, participants were informed via the computer screen that all collected data would be analyzed anonymously, and their rights to withdraw at any time (without consequences for payment) were explained. By pressing a key, they agreed to participate in the experiment.

After this, they completed the need to belong scale (adapted from Leary, Kelly, Cottrell, & Schreindorfer, 2007, 10 items, 7-point scale, α = .87, see Appendix A.6 for the items on this scale). This scale was included to check for differences in participants' need to belong before the exclusion manipulation.

Figure 2.1: Screenshot of the Cyberball game with the participant (Me) at the bottom and the two computerized players (Herman and Jolanda) at the top. In the screenshot, player Jolanda is throwing the ball to player Herman.

To make participants feel socially included or excluded, they played Cyberball (Williams & Jarvis, 2006). This game is often used to experimentally induce people to feel socially excluded (e.g., Williams, Cheung, & Choi, 2000; Zadro, Williams, & Richardson, 2004; Williams & Jarvis, 2006; van Beest & Williams, 2006; Bernstein & Claypool, 2012). Cyberball is a virtual ball tossing game with two computerized players. The game instructions were the following: 'when the ball is tossed to you, you need to click on one of the other two players to toss the ball to that player'. A screenshot of the game is presented in Figure 2.1. Participants in the social inclusion condition received the ball roughly one third of the throws. Participants in the social exclusion condition received the ball once or twice in the beginning of the game, after which the computerized players only tossed the ball to each other during the remainder of the game. As a cover story, participants were told that this game was necessary for practicing mental visualization, which was stated to be needed in the washing machine task (which will be explained in the next paragraph). Participants in the control condition did not play Cyberball but instead continued with the next part of the experiment.

Figure 2.2: The artificial agent Femke used in Studies 1 and 2.

After playing Cyberball, participants were introduced to artificial agent Femke (see Figure 2.2) that would provide feedback about choices they made during the washing machine task. The agent was presented on a second monitor and was programmed to respond to every choice participants made in the task. Next, participants started with the washing machine task as developed by McCalley and Midden (2002) and used in earlier PT research (e.g., Midden & Ham, 2008; Vossen et al., 2009; Vossen, Ham, & Midden, 2010; Midden & Ham, 2012). In this task, a simulated washing machine interface (see Figure 2.3) was presented on the screen and participants were asked to complete ten laundry tasks (e.g., 'Wash four dirty jeans.'). The ten trials were randomly selected from a set of 23 different laundry tasks (see Appendix B for a list of these tasks).

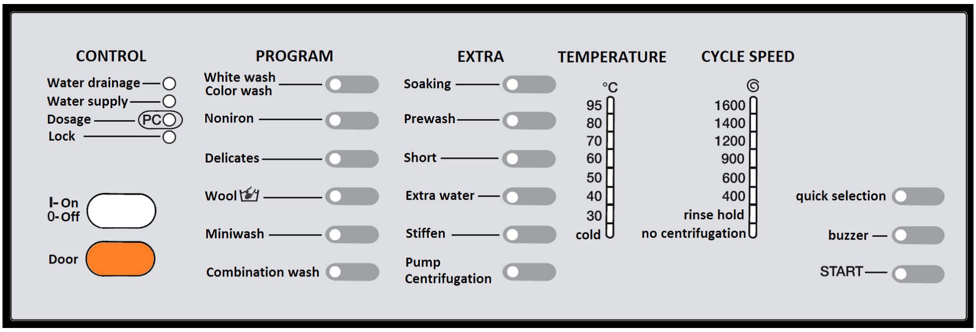


Figure 2.3: Screenshot of the simulated washing machine interface.

During the setting of the simulated washing machine, participants received feedback from the artificial agent. This feedback was based on the energy use with the chosen settings and had six levels, three of which positive (i.e. 'Fantastic!', 'Very good!', and 'Well done!') and three negative (i.e. 'Terrible!', 'Very bad!', and 'That's bad!').

The main dependent variable was created as follows. First, the energy use with the chosen settings was standardized to be able to compare between the different laundry tasks. Next, the difference on this standardized value between every setting and the previous setting (caused by an action performed by the participant, e.g., increasing the washing temperature) was calculated. This value shows the immediate effect of the agent's feedback on participants' energy use. A positive energy use indicated that a participant used more energy than in the previous setting, and a negative energy use indicated that the participant used less energy compared with the previous setting. When participants first chose a program for the simulated washing machine, all other settings (temperature and cycle speed) were automatically set to the highest possible value. By changing these settings, participants could try to use less energy. Because of this, participants' average energy use was most likely to be negative. For ease of interpretation, energy use was converted into Energy-conservation (by multiplying it by -1) to represent the amount of energy that was saved.

After performing the washing machine task, participants completed two self-report anthropomorphism questionnaires to measure perceived human-likeness of the artificial agent. For this, two questionnaires that are often included in research that investigates effects of anthropomorphism were used. These questionnaires were the 5-item anthropomorphism scale, adapted from the Godspeed questionnaire as developed by Bartneck et al. (2009) (7-point scale, $\alpha = .76$, see Appendix A.1 for the items on this scale), and a 7-item questionnaire, adapted from Waytz, Morewedge, et al. (2010) (7-point scale, $\alpha = .86$, see Appendix A.2 for the items on this scale). The responses on these questionnaires were averaged and used as self-report measurements of anthropomorphism. These questionnaires will be referred to as the Godspeed-instrument and the Waytz-instrument respectively.

Next, the effectiveness of the manipulation was checked by asking participants to estimate the percentage of balls they had received during the Cyberball game (referred to as Ball-percentage), followed by 13 questions about participants' evaluation of the game, based on Zadro et al. (2004) ($\alpha = .93$, referred to as Game-evaluation, see Appendix A.3 for the items on this scale). Participants in the control condition skipped these manipulation checks.

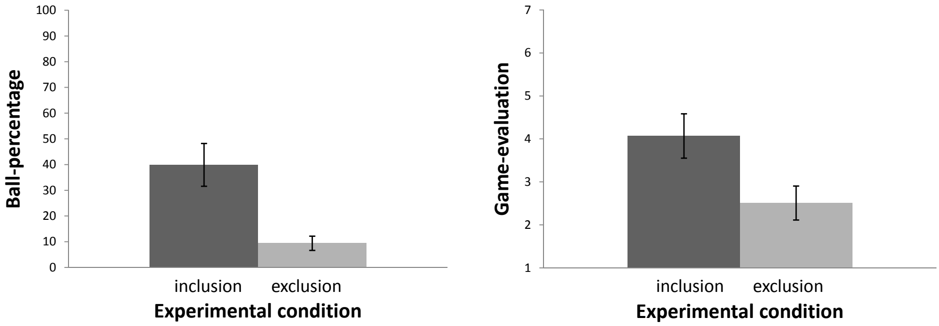
Participants' responses to the artificial agent could be influenced by other factors than its perceived human-likeness. For example, when the agent is perceived as intelligent or knowledgeable, people may adapt their behavior more to its feedback than when it is perceived as unintelligent or ignorant. For this reason, perceived intelligence of the artificial agent (adapted from Bartneck et al., 2009, 5 items, 7-point scale, $\alpha = .87$, see Appendix A.4 for the items on this scale) and perceived agent knowledge (adapted from Powers & Kiesler, 2006, 8 items, 7-point scale, $\alpha = .87$, see Appendix A.5 for the items on this scale) were also measured.

At the end of the session, participants indicated their age and gender. After this, they were debriefed about the goal of the social exclusion manipulation and the feedback provided by the artificial agent, were paid and thanked for their contribution.

2.2.2 Results

Manipulation check

To check whether the social exclusion manipulation was successful, results on the two manipulation checks (Ball-percentage and Game-evaluation) were submitted to independent samples t-tests with the two Cyberball conditions as groups. Results showed a significant difference between the groups on Ball-percentage, $t(41) = 8.15$, $p < .001$, $r^2 = .72$. Participants in the social inclusion condition ($M = 39.90$, $SD = 16.62$) reported a higher percentage of balls received during the game than participants in the social exclusion condition ($M = 9.41$, $SD = 5.56$). Participants in the social inclusion condition ($M = 4.07$, $SD = 1.03$) also evaluated the game more positively than participants in the social exclusion condition ($M = 2.51$, $SD = 0.79$), $t(41) = 5.59$, $p < .001$, $r^2 = .42$. These effects are visualized in Figure 2.4.



(a) (b)

Figure 2.4: Visualization of (a) ball-percentage and (b) game-evaluation per Cyberball condition in Study 1. Whiskers represent 95% error bars.

Also, no difference was found between the three experimental conditions on participants' need to belong, $F(2,\ 62) = 0.04$, $p = .96$. A follow-up pairwise comparison showed no differences between the social exclusion and inclusion condition, $t(60) = 0.27$, $p = .79$. Thus, participants' need to belong at the start of the experiment was similar in all experimental conditions.

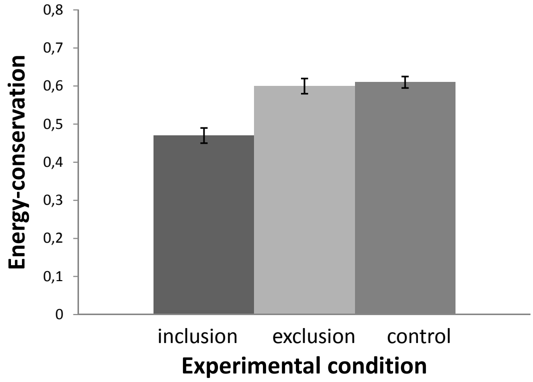
Energy conservation

To test whether social exclusion and gender influenced susceptibility to the social feedback, Energy-conservation was submitted to a Linear Mixed Model with a 3 (Condition: control vs. inclusion vs. exclusion) X 2 (Participant-gender: male vs. female) between subjects design with the specific laundry task as random factor. This analysis showed a significant main effect of Condition, $F(2,\ 2475.37) = 3.70$, $p = .03$, $\omega^2 < .001$[1]. Participants in the control condition ($EMM = 0.61$, $SE = 0.03$) saved the most energy, followed by those who were socially excluded ($EMM = 0.60$, $SE = 0.04$), and those who were socially included ($EMM = 0.47$,

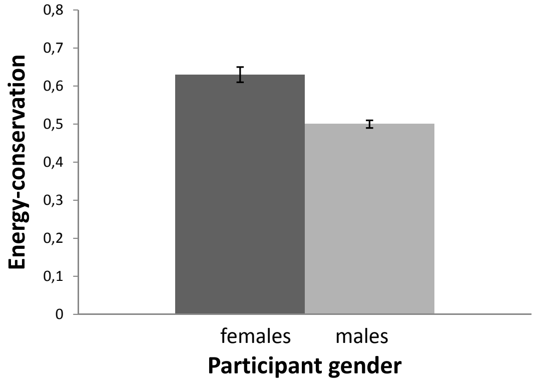
[1] ω^2's were calculated using the method obtained from Xu (2003)

$SE = 0.04$). This effect is visualized in Figure 2.5a.

Follow-up pair-wise comparisons revealed significant differences between the social exclusion and inclusion condition ($p = .03$, $\omega^2 < .001$) and between the social inclusion and control condition ($p = .01$, $\omega^2 < .001$), but not between the social exclusion and control condition ($p = .86$).



(a)



(b)

Figure 2.5: Visualization of energy-conservation per (a) experimental condition and (b) participant gender in Study 1. Whiskers represent 95% confidence intervals.

Additionally, a significant main effect of Participant-gender was found, $F(1, 2580.66) = 8.45$, $p < .01$, $\omega^2 = .03$. More specifically, female participants ($EMM = 0.63$, $SE = 0.04$) saved more energy than male participants (*EMM* = 0.50, *SE* = 0.02). This effect is visualized in Figure 2.5b. Results provided no evidence for an interaction between Condition and Participant-gender, $F(2, 2475.37) = 0.18$, $p = .83$.

Anthropomorphism

To check whether social exclusion had an effect on self-reported anthropomorphism, the averaged scores on the Godspeed- and Waytz-instruments were submitted to an analysis of variance (ANOVA) with the three experimental conditions as groups. Results showed no significant effect of experimental condition on either the Godspeed-instrument ($F(2, 62) = 0.14$, $p = .87$) or the Waytz-instrument ($F(2, 62) = 1.33$, $p = .27$). Also, pairwise comparisons showed no difference between the social exclusion and inclusion conditions for both the Godspeed-instrument ($t(60) = 0.48$, $p = .64$) and the Waytz-instrument ($t(60) = 1.59$, $p = .12$). Thus, social exclusion did not influence participants' self-reported anthropomorphism of the artificial agent.

Interestingly, a significant positive correlation was found between participants' need to belong and self-reported anthropomorphism on both the Godspeed-instrument ($r(63) = .25$, $p = .05$) and the Waytz-instrument ($r(63) = .33$, $p < .01$).

No effect of experimental condition was found on either perceived intelligence ($F(2, 62) = 0.68$, $p = .51$) or perceived agent knowledge ($F(2, 62) = 1.21$, $p = .31$). Also, pairwise comparisons showed no difference between the social exclusion and inclusion conditions on both perceived intelligence ($t(60) = 0.50$, $p = .62$) and perceived agent knowledge ($t(60) = 1.19$, $p = .24$). Thus, social exclusion did not influence the extent to which the artificial agent was perceived as intelligent or knowledgeable.

To investigate the effects of participant gender on self-reported anthropomorphism, the averaged scores of male and female participants on the Godspeed- and the Waytz-instruments were submitted to independent samples t-tests with males and females as groups. Results showed no significant difference on the Waytz-instrument ($t(39.07) = 1.05$, $p = .30$), but they did show a significant difference on the Godspeed-instrument $t(61) = 2.39$, $p = .02$, $r^2 = .09$. More specifically, female participants ($M = 3.62$, $SD = 0.97$) reported higher levels of anthropomorphism than male participants ($M = 3.02$, $SD = 0.97$). This effect is visualized in Figure 2.6.

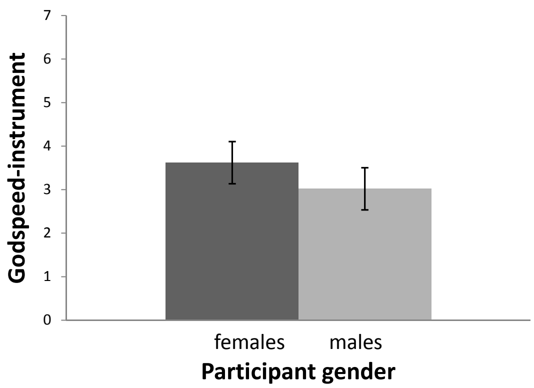


Figure 2.6: Visualization of averaged anthropomorphism scores on the Godspeed-instrument per participant gender in Study 1. Whiskers represent 95% error bars.

2.2.3 Discussion

This study was designed to investigate effects of social exclusion on susceptibility to persuasion by an artificial agent. Results showed that participants who were experimentally induced to feel socially excluded showed more behavior change as a result of the social feedback from an artificial agent compared to those who were socially included. These results supported the main hypothesis. In addition, a control condition was included to investigate whether social exclusion would increase susceptibility or social inclusion would decrease susceptibility to social influence. Pair-wise comparisons between this control condition and the social exclusion and inclusion conditions revealed a difference between the control condition and the social inclusion condition, but not between the control condition and the social exclusion condition.

One possible explanation for this finding may be that playing a game in which a person is socially *included* may decrease that person's need for social contact compared to not playing the game. As a result, this person might be less susceptible to persuasion by an artificial agent. In contrast, being socially *excluded* in the game may not affect a person's need for social contact compared to not playing the game. Consequently, people who were socially excluded in the game might have had a need for social contact that was similar to that of those who did not play the game at all. Future studies could be designed to investigate the direction of the effect more extensively, that is whether social exclusion *increases* or social inclusion *decreases* a person's susceptibility to persuasion. While this is an interesting question, it is outside of the scope of the current chapter.

Results of the current study showed no differences on either perceived intelligence or perceived human-likeness of the artificial agent between any of the conditions. These findings suggest that both those concepts do not explain the effect of social exclusion on susceptibility to persuasion by the artificial agent. However, moderate positive correlations were found between anthropomorphism and need to belong, showing that participants with a higher need to belong attributed more human-likeness to the artificial agent. This latter finding is in line with earlier ones (e.g., Epley, Akalis, et al., 2008; Waytz, Cacioppo, & Epley, 2010; Eyssel & Reich, 2013). However, the hypothesis that social exclusion would lead to higher attributions of human-likeness was not confirmed, even though social exclusion did make participants adapt their behavior more in the washing machine task than social inclusion. One possible explanation for this is that the social exclusion manipulation did influence participants' (less controlled) behavior in the washing machine task, but did not influence their (more controlled) responses on the self-report questionnaires. This issue will be discussed more in the general discussion of this chapter.

In addition to the effect of social exclusion on behavior change, an effect of participant gender was found. That is, female participants appeared to be more susceptible to the social feedback than male participants. This result supports earlier findings that susceptibility to social influence differs between males and females (e.g., Eagly, 1978; Eagly & Carli, 1981). Furthermore, such a gender effect has also been shown in the context of persuasive support strategies (Orji, 2014). In contrast to earlier findings, however, no interaction between participant gender and social exclusion was found. Williams and Sommer (1997) reported that females worked harder in a collective task and that only ostracized female participants showed more engagement in the task. However, they did not report means of the different groups, nor did they show any results on the non-significant contrasts (Williams & Sommer, 1997). Therefore, it is difficult to determine the strength and relevance of their reported interaction. Moreover, findings in the current study do not support such an interaction.

The gender effect in the current study could be interpreted in another way as well. All participants interacted with a female artificial agent. Thus, female participants interacted with an artificial agent of the *same* gender, whereas male participants interacted with an artificial agent of the *opposite* gender. Research in human-human interactions showed that gender differences had a stronger influence in same-gender interactions compared to mixed-gender interactions (Carli, 2013). Artificial agents have also been shown to have a greater influence on people's attitudes when those agents have the same gender as the participant (Guadagno, Blascovich, Bailenson, & Mccall, 2007). This gender-matching effect could have caused the effect found in the current study. A second study was designed with two purposes: to replicate the exclusion effect, and to further investigate the gender effect found in the current study.

2.3 Study 2

In Study 1, effects of social exclusion on susceptibility to social influence coming from an artificial agent were investigated. Results showed that socially excluded people were more susceptible to the social feedback than socially included people. Furthermore, an effect of participant gender was found. That is, female participants appeared to be influenced more by the social feedback than male participants. This effect could be explained in two different ways. First, participants could show a greater attitude change when persuaded by a same-gender artificial agent. Earlier work has shown that people are more susceptible to social influence in same-gender interactions than in mixed-gender interactions (Guadagno et al., 2007; Carli, 2013). Second, women in general could be more susceptible to social influence than men. Results of a review showed that women are more persuadable than men in group pressure situations (Eagly & Carli, 1981).

The main aim of the current study was to replicate the social exclusion effect found in Study 1. The main hypothesis was that social exclusion would induce behavior change as a result of social feedback from an artificial agent. In addition, the current study explored whether the gender effect found in Study 1 is more likely to be caused by the gender of the participant or by the gender-matching between the participant and the artificial agent.

2.3.1 Method

Participants and design

Eighty-nine participants (60 males and 29 females; M_{age} = 24.0, SD_{age} = 10.1, Range = 18 to 64) were randomly assigned to one of four experimental conditions of a 2 (Cyberball: inclusion vs. exclusion) X 2

(Gender-matching: same-gender vs. mixed-gender) between-subjects design. The Cyberball conditions were identical to those in Study 1, and were designed to make participants feel either socially included ($n = 44$) or socially excluded ($n = 45$). The Gender-matching conditions (same-gender $n = 49$, mixed-gender $n = 40$) were created by having participants interact with either a male or a female artificial agent. An overview of the number of male and female participants in each experimental condition is presented in Table 2.1. The experiment lasted 30 minutes for which participants were paid €5.

Table 2.1: Number of male and female participants per experimental condition in Study 2.

	Inclusion		Exclusion	
	Male	Female	Male	Female
Same-gender	17	7	17	8
Mixed-gender	12	8	14	6

Materials and procedure

At the start of the experiment, participants were informed via the computer screen that all collected data would be analyzed anonymously, and their rights to withdraw at any time (without consequences for payment) were explained. By pressing a key, they agreed to participate in the experiment.

After this, they completed the need to belong scale (10 items, 7-point scale, $\alpha = .79$). As in Study 1, this scale was included to check for differences in participants' need to belong before the exclusion manipulation.

To make participants feel socially included or excluded, they played the Cyberball game (Williams & Jarvis, 2006) as used in Study 1. As a cover story, participants were told that this game was necessary for practicing mental visualization, which was stated to be needed for the washing machine task.

Figure 2.7: The artificial agent Max used in Study 2.

After playing Cyberball, participants were introduced to one of two artificial agents (either Max, see Figure 2.7, or Femke[2]), depending on the gender-matching condition they were in. The agent provided social feedback on participants' choices during the washing machine task. The agent was presented on a second monitor and was programmed to respond to every choice participants made in the task. Next, participants started with the main experimental task: the washing machine task as explained in Study 1. For this study, ten of the 23 laundry trials were selected (see Appendix B), so that all participants would perform the same tasks in random order throughout the experiment. Both the feedback levels and the main dependent variable (Energy-conservation) were identical to those in Study 1.

[2]To test for differences in the perceived human-likeness of these artificial agents, both of them were included in a pretest in which a group of 70 participants (36 males and 34 females; M_{age} = 26.8, SD_{age} = 12.8, Range = 14 to 61) each rated two of four different agents on anthropomorphism using the Godspeed-instrument (5 items, α = .83) and the Waytz-instrument (7 items, α = .82). The two artificial agents to be rated were selected randomly, making 39 participants rate Max and 34 participants rate Femke. When comparing these two agents on their perceived human-likeness, no differences were found on either the Godspeed-instrument, $F(1, 72) = 0.06$, $p = .81$, or the Waytz-instrument, $F(1, 72) = 0.46$, $p = .50$.

After performing the washing machine task, participants completed two self-report anthropomorphism questionnaires to measure perceived human-likeness of the artificial agent. For this, the Godspeed-instrument (5 items, 7-point scale, $\alpha = .75$) and the Waytz-instrument (7 items, 4-point scale, $\alpha = .84$) were used.

Next, the effectiveness of the manipulation was checked by asking participants to estimate the percentage of balls they had received during the Cyberball game (referred to as Ball-percentage), followed by 14 questions about their evaluation of the game, based on Zadro et al. (2004) ($\alpha = .92$, referred to as Game-evaluation).

After this, participants' general evaluations of the artificial agents were measured. This questionnaire consisted of 9 items ($\alpha = .86$, see Appendix A.7 for the items on this scale). No effects were found on this scale, showing that the agents were not evaluated differently.

At the end of the session, participants indicated their age and gender. After this, they were debriefed about the goal of the social exclusion manipulation and the feedback provided by the artificial agent, were paid and thanked for their contribution.

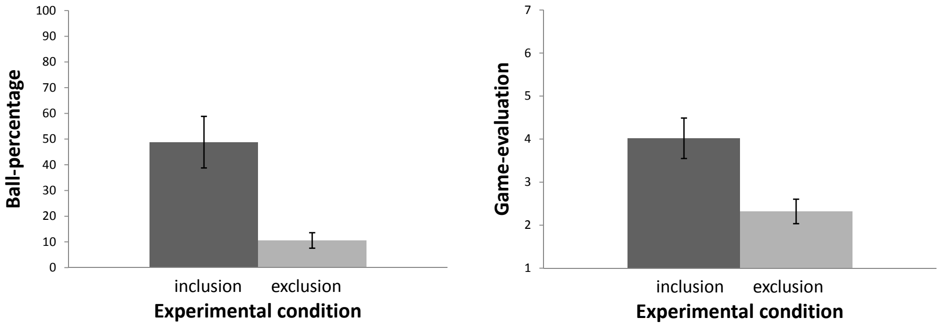
2.3.2 Results

Manipulation check

To check whether the manipulation was successful, results on the two manipulation checks (Ball-percentage and Game-evaluation) were submitted to independent samples t-tests with the two Cyberball conditions as groups[3]. Results showed a significant difference between the groups on Ball-percentage, $t(65) = 11.68$, $p < .001$, $r^2 = .71$. Participants in

[3]Due to a program error, data from one experimental condition (social exclusion with the female artificial agent) were not stored properly. Therefore, manipulation checks could only be performed on three of the four conditions.

the social inclusion condition (M = 48.82, SD = 20.08) reported having received a higher percentage of the balls during the game than participants in the social exclusion condition (M = 10.57, SD = 6.00). Also, participants in the social inclusion condition (M = 4.02, SD = 0.94) evaluated the game more positively than participants in the social exclusion condition (M = 2.32, SD = 0.57), $t(66.58) = 9.31$, $p < .001$, $r^2 = .57$. These effects are visualized in Figure 2.8.



(a) (b)

Figure 2.8: Visualization of (a) ball-percentage and (b) game-evaluation per Cyberball condition in Study 2. Whiskers represent 95% error bars.

This time, a significant difference was found between the social exclusion and inclusion conditions on participants' need to belong, $F(1, 88) = 5.43$, $p = .02$. Participants who were going to be socially excluded (M = 4.35, SD = 0.78) had a lower need to belong at the start of the experiment than participants who were going to be socially included (M = 4.76, SD = 0.88). For this reason, need to belong will be included as random factor in the energy conservation analyses.

Energy conservation

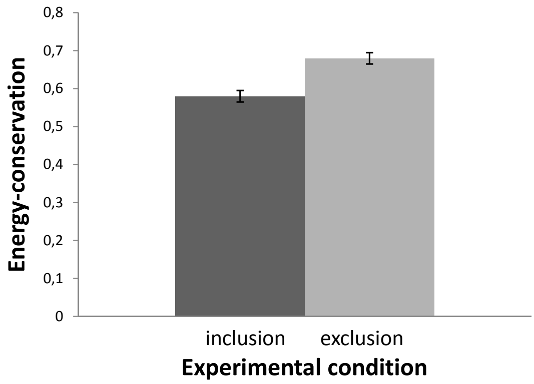
To test whether social exclusion, participant gender, and gender-matching influenced susceptibility to the social feedback, Energy-conservation was submitted to a Linear Mixed Model with a 2 (Cyberball: inclusion vs.

exclusion) X 2 (Participant-gender: male vs. female) X 2 (Gender-matching: same-gender vs. mixed-gender) design with the specific laundry task and need to belong as random factors. The analysis showed a significant main effect of Cyberball, $F(1,\ 2742.52) = 6.58$, $p = .01$, $\omega^2 < .001$. Participants who were excluded (*EMM* = 0.68, *SE* = 0.03) saved more energy than participants who were included (*EMM* = 0.58, *SE* = 0.03). This effect is visualized in Figure 2.9a.

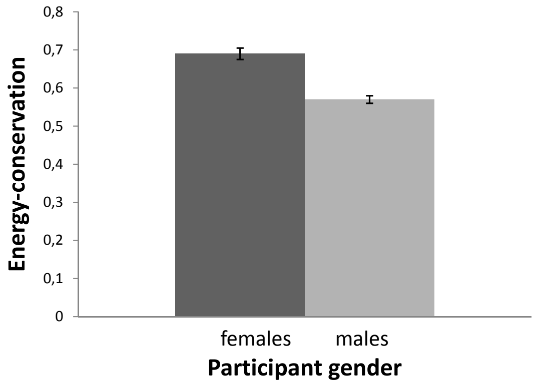
Additionally, this analysis revealed a main effect of Participant-gender, $F(1,\ 2587.14) = 9.15$, $p < .01$, $\omega^2 = .002$. More specifically, female participants (*EMM* = 0.69, *SE* = 0.03) saved more energy than male participants (*EMM* = 0.57, *SE* = 0.02). This effect is visualized in Figure 2.9b. Results provided no evidence for an effect of Gender-matching ($F(1,\ 2391.62) = 0.02$, $p = .89$).

Interestingly, an interaction was found between Cyberball and Gender-matching, $F(1,\ 2391.62) = 3.71$, $p = .05$, $\omega^2 < .001$. More specifically, participants in the mixed-gender condition who were socially excluded (*EMM* = 0.72, *SE* = 0.05) saved more energy than those who were socially included (*EMM* = 0.54, *SE* = 0.03), whereas in the same-gender condition, there was no difference in energy use between participants who were excluded (*EMM* = 0.64, *SE* = 0.04) and those who were included (*EMM* = 0.61, *SE* = 0.04). This interaction effect is visualized in Figure 2.9c.

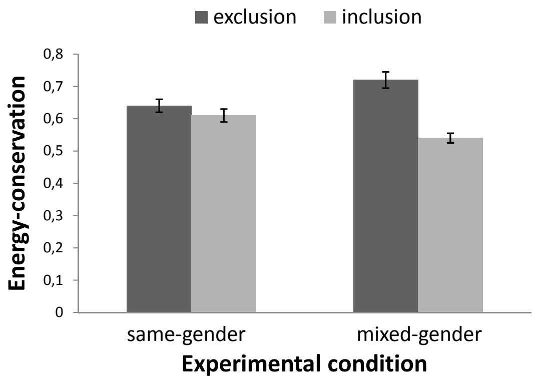
Follow-up pair-wise comparisons revealed significant differences between the excluded and included participants in the mixed-gender condition ($p < .01$, $\omega^2 < .001$), but not in the same-gender condition ($p = .54$).



(a)



(b)



(c)

Figure 2.9: Visualization of energy-conservation (a) per Cyberball condition, (b) per participant gender, and (c) the interaction between Cyberball and gender-matching in Study 2. Whiskers represent 95% confidence intervals.

Anthropomorphism

To check whether social exclusion had an effect on self-reported anthropomorphism, the averaged scores on both the Godspeed- and the Waytz-instruments were submitted to independent samples t-tests with the two Cyberball conditions as groups. Results showed no significant difference on the Godspeed-instrument ($t(87) = 0.70$, $p = .49$). A significant difference between the Cyberball conditions was found on the Waytz-instrument, $t(87) = 2.18$, $p = .03$, $r^2 = .05$. In contrast to the expectation, participants in the social inclusion condition ($M = 2.03$, $SD = 0.74$) reported higher levels of anthropomorphism than participants in the social exclusion condition ($M = 1.71$, $SD = 0.67$). This effect is visualized in Figure 2.10.

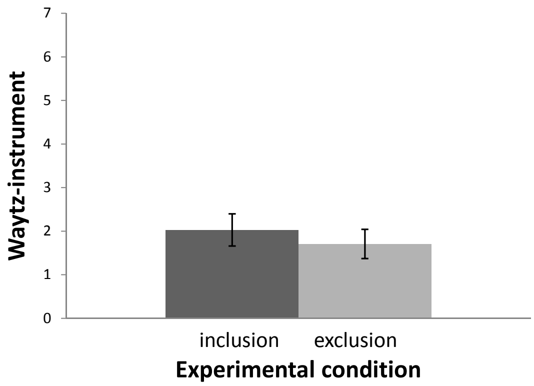
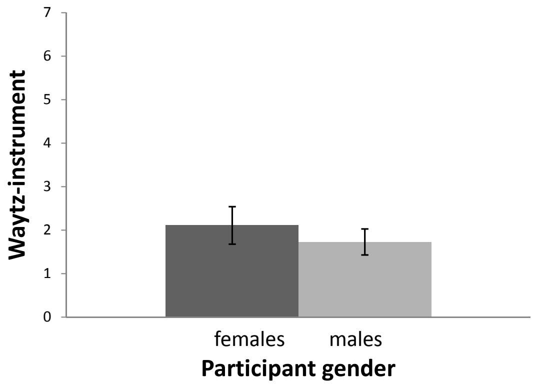
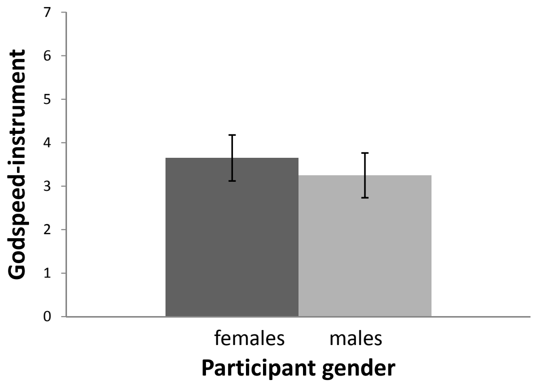


Figure 2.10: Visualization of averaged scores on the Waytz-instrument per Cyberball condition in Study 2. Whiskers represent 95% error bars.

This time, no significant correlation was found between self-reported anthropomorphism and participants' need to belong for both the Godspeed-instrument ($r(89) = .11$, $p = .30$) and the Waytz-instrument ($r(89) < .01$ $p = .96$).



(a)



(b)

Figure 2.11: Visualization of averaged scores per participant gender on (a) the Godspeed-instrument and (b) the Waytz-instrument in Study 2. Whiskers represent 95% error bars.

To investigate the effects of Participant gender on self-reported anthropomorphism, the averaged scores of male and female participants on the Godspeed- and the Waytz-instruments were submitted to independent samples t-tests with males and females as groups. Results showed a significant difference on the Waytz-instrument, $t(42.49) = 2.36$, $p = .02$, $r^2 = .07$. More specifically, female participants ($M = 2.15$, $SD = 0.85$) reported higher levels of anthropomorphism than male participants ($M = 1.73$, $SD = 0.61$). Results also showed a significant difference on the Godspeed-instrument $t(87) = 2.29$, $p = .03$, $r^2 = .06$. Again, female participants (M = 3.74, SD = 0.95) reported higher levels of anthropomorphism than male participants (M = 3.23, SD = 1.02). These effects are visualized in Figure 2.11.

2.3.3 Discussion

The main aim of the current study was to investigate the effect of social exclusion on susceptibility to social feedback from an artificial agent. Results showed that people who were experimentally induced to feel socially excluded showed more behavior change as a result from the social feedback compared to those who were socially included. These results confirmed the main hypothesis and replicated the findings of Study 1. The current study also explored whether female participants were more susceptible to social feedback than male participants, or whether participants were more susceptible to social feedback in same-gender interactions compared to mixed-gender interactions. Results showed that participants' gender influenced their behavior, whereas gender-matching did not have such an influence. However, an interaction was found between social exclusion and gender-matching. More specifically, in the mixed-gender condition, socially excluded participants saved more energy than socially included participants, whereas in the same-gender condition there was no difference between social exclusion and inclusion. Thus, in contrast to earlier findings that showed that people are generally more susceptible to social influence in same-gender interactions (see e.g., Guadagno et al., 2007; Carli, 2013), being socially excluded makes them more susceptible to persuasion by an artificial agent in mixed-gender interactions.

Similar results were found by Harmon-Jones, Peterson, and Harris (2009). In their studies, participants were socially excluded by a person of the opposite gender or by a person of the same gender during the Cyberball game. Results showed that being socially excluded by a person of the opposite gender leads to stronger feelings of jealousy (Harmon-Jones et al., 2009). Research in the field of persuasive robotics also showed that people rated a robot as more credible and were more likely to donate money to the robot when it was of the opposite gender than when it was of the same gender as the participant (Siegel et al., 2009). Findings in the current chapter showed that susceptibility to social influence after being socially excluded is stronger in mixed-gender interactions than in same-gender interactions. Thus, the user's emotional state and the agent's gender together determine that agent's persuasiveness. This highlights the importance of including both the users' emotional state and the agents' gender in the design of social interactions with artificial agents.

In line with results of Study 1, female participants in the current study attributed more human-likeness to the artificial agent than male ones, and they also saved more energy than males. These results were in line with the expected relation between anthropomorphism and susceptibility to an artificial agent's social feedback. Social exclusion was found to *increase* susceptibility to social feedback, but it *decreased* reported levels of anthropomorphism. More specifically, socially excluded participants saved more energy than socially included participants, but they also reported lower levels of anthropomorphism on the Waytz-instrument than socially included participants. This latter finding was opposite of what was expected.

One explanation for the finding that socially excluded participants attributed less human-likeness to the artificial agent than socially included participants could be that the artificial agent's socialness did not meet participants' expectations. Participants who were socially excluded were expected to have experienced a higher need for social contact than participants who were socially included, and were thus expected to be more attentive to social cues. These social cues were shown by the artificial agent. However, the artificial agent was programmed to respond every time a participant changed a setting on the washing machine interface. This could have made the artificial agent come across as static and machine-like, instead of dynamic and human-like, especially for the participants who were socially excluded and thus expected to be more attentive to social cues. When they were asked to rate the agent on its human-likeness, these participants could have evaluated the agent as less rather than more human-like. An alternative explanation for the unexpected finding on reported levels of anthropomorphism is that participants who were socially excluded disliked the experiment more than those who were socially included. As a result, this could have reduced the social desirability bias (see e.g., Edwards, 1957; Fisher, 1993) which prompts people to go along with the socialness of the artificial agent. Future research could be designed to investigate people's evaluations of artificial agents that do not meet their expectations with respect to their socialness, or whether social desirability could influence people's attributions of human-likeness to artificial agents.

Another explanation is that participants who were going to be socially excluded started the experiment with a lower need to belong than those who were going to be socially included. The social exclusion manipulation could have influenced participants' need to belong to the extent that it became equally high in both conditions. This would suggest that other processes than anthropomorphism caused the difference in participants' susceptibility to the social feedback from the artificial agent, also because no significant correlation between need to belong and anthropomorphism was found in the current study. This issue will be discussed more in the general discussion of this chapter.

2.4 General discussion

The current research was designed to investigate effects of social connectedness as a psychological determinant of anthropomorphism on susceptibility to social influence by an artificial agent. Earlier research on social connectedness has shown that social exclusion increases both attentiveness to social cues (Pickett et al., 2004) and attributions of human-likeness to artificial agents (Epley, Waytz, et al., 2008). Furthermore, it has been demonstrated that perceived human-likeness of agents influences people's behavioral intentions to participate in energy conservation (Ahn et al., 2014). We hypothesized that people who are socially excluded would attribute more human-likeness and would be more susceptible to persuasion by an artificial agent than people who are socially included. These hypotheses were explored in two studies. Results of Study 1 showed that social exclusion caused more behavior change in a simulated washing machine task than social inclusion. This finding supported the main hypothesis and was replicated in Study 2. No support was found for the hypothesis that social exclusion also increases perceived human-likeness of the artificial agents. In Study 1, no difference on perceived human-likeness was found between the social inclusion and exclusion groups. In Study 2, social inclusion even made people attribute higher levels of anthropomorphism to the artificial agents than social exclusion.

I speculated that the way in which the artificial agent provided feedback could have made it come across as less social, especially for people who were more attentive to social cues. More specifically, every time a participant changed a setting on the washing machine interface, the agent reacted by uttering a word or a phrase that reflected the current energy use. This could be experienced as machine-like rather than human-like, causing people to attribute *less* human-likeness to the agent *when explicitly asked about it.* People's automatic responses to technology are shown to be predominantly social (for an overview, see Reeves & Nass, 1996), but when explicitly asked about it, people actually denied that they actually showed such social behavior (see Nass & Moon, 2000). This finding is in line with theories about automatic versus controlled processing, which argue that two different kinds of processes could cause different responses (e.g., Schneider & Shiffrin, 1977; Evans, 2003). When people explicitly evaluate an artificial agent on its human-like characteristics, they use controlled processes, whereas their behavior in the washing machine task was more likely to be caused by automatic processes. These findings suggest a need for measuring people's attributions of human-like characteristics to technology in a less obtrusive way. In particular, participants were asked to reflect on the artificial agents' consciousness and free will, human-like characteristics that might be difficult to attribute to such an artificial agent. Consequently, the used measuring instruments could have been insufficient for measuring the effects of the experimental manipulation on perceived human-likeness. For this reason, different measuring instruments for anthropomorphism will be used in Chapter 3, and this issue will be extensively discussed in Chapter 4.

Another possible explanation for not finding the expected effects on perceived human-likeness may be that the effect of social exclusion on persuasion is caused by a different process than anthropomorphism. In Study 1, perceived intelligence of the agent was measured, but results showed no differences between the experimental conditions. There could be other personality characteristics that may explain the effect of social exclusion on susceptibility to persuasion as well. For example, self-esteem has been shown to correlate with susceptibility to social influence (e.g., Janis, 1954; Berkowitz & Lundy, 1957; Bearden, Netemeyer, & Teel, 1989; Allen, Chango, Szwedo, Schad, & Marston, 2012). Janis (1954) concluded that people with low self-esteem tend to be more susceptible to social influence than others. Likewise, Berkowitz and Lundy (1957) found that people who are low in interpersonal confidence are more susceptible to peer influence than those high in interpersonal confidence. When people are socially excluded, this may temporarily decrease their self-esteem, causing them to be more sensitive to social influence by the artificial agent. In the studies presented in this chapter, self-esteem was not measured. More research is needed to investigate whether such personality traits could indeed moderate effects of social exclusion.

In addition to effects of social exclusion, effects of participants' gender on their behavior in the washing machine task were found. In both Study 1 and Study 2, female participants were more susceptible to social feedback than male ones. This finding was in line with earlier findings on gender effects in the context of social exclusion (Williams & Sommer, 1997) and persuasion (Orji, 2014). Additionally, an interaction was found between social exclusion and gender-matching in Study 2. On the one hand, this interaction contradicted earlier findings on persuasion in mixed-gender interactions (e.g., Guadagno et al., 2007). On the other hand, it supported findings on social exclusion in mixed-gender interactions (Harmon-Jones et al., 2009). Gender effects in social influence have to my knowledge not been investigated in the context of psychological states like feelings of belonging, and should be taken into account in future research. This future research could be designed to investigate the role of participant gender and gender-matching in interactions with artificial agents.

2.4.1 Limitations and future research

Participants in the studies in the current chapter performed a simulated washing machine task on a computer screen in a controlled lab environment. Although this procedure has been validated in field research through previous studies (e.g., McCalley, Kaiser, Midden, Keser, & Teunissen, 2006), people's behavior in the lab may be considerably different from that in a real-life situation (e.g., Levitt & List, 2007). Therefore, future research should also investigate effects of social exclusion in field settings.

Additionally, participants in the studies in the current chapter were induced to feel socially excluded, but it was not assessed to what extent they actually experienced the negative consequences of being socially excluded. They were asked how they felt during the game, but the consequences this may have had on their feelings during the remainder of the experiment are unknown. Future research could be designed to investigate whether people actually feel more lonely after playing Cyberball, and how long this feeling lasts.

2.4.2 Conclusions

Overall, results presented in the current chapter are promising and adding to the body of literature in the domain of artificial agents as PT. Results showed that an artificial agent's persuasiveness can be influenced by a user's psychological state, suggesting that adaptive properties of artificial agents that could take these psychological states into account could contribute to the effectiveness of those agents (for an example of the effectiveness of adaptive PT, see Kaptein & van Halteren, 2013).

The possibility to adapt to a user's psychological or emotional state could be an important design challenge for future development of artificial agents as PT. Such psychological or emotional states could influence a person's tendency to anthropomorphize artificial agents, which in turn could increase their susceptibility to persuasive messages coming from those agents.

The studies in the current chapter investigated the role of a *psychological* determinant of anthropomorphism on persuasion by artificial agents. The studies in the next chapter will investigate the role of a *technological* determinant of anthropomorphism: consistency of social cues.

"In making a speech one must study three points: first, the means of producing persuasion; second, the language; third the proper arrangement of the various parts of the speech."

Aristotle (384BC – 322BC)

3

Ambiguous agents

The influence of consistency of an artificial agent's social cues on emotion recognition, recall, and persuasiveness

---

This chapter is largely based on:

Ruijten, P. A. M., Midden, C. J. H, and Ham, J. (2014). *Ambiguous agents: On the influence of consistency of an artificial agent's social cues on emotion recognition, recall, and persuasiveness.* Manuscript submitted for publication.

Ruijten, P. A. M., Midden, C. J. H., and Ham, J. (2013). I didn't know that virtual agent was angry at me: Investigating effects of gaze direction on emotion recognition and evaluation. In: S. Berkovsky & J. Freyne (Eds.): *PERSUASIVE 2013, LNCS 7822* (pp. 192-197). Springer-Verslag Berlin Heidelberg 2013.

This chapter explores the relation between a *technological* determinant of anthropomorphism, consistency of social cues, and persuasion by an artificial agent. Including (minimal) social cues in Persuasive Technology (PT) increases the probability that people attribute human-like characteristics to that technology, which in turn can make that technology more persuasive (see e.g., Nass et al., 1993). PT in the social actor role can be equipped with a variety of social cues to create opportunities for applying social influence strategies (for an overview, see Fogg, 2003). However, multiple social cues may not always be perceived as being consistent, which could decrease their perceived human-likeness and their persuasiveness. In the current chapter, I investigate the relation between consistency of social cues and persuasion by an artificial agent.

3.1 General introduction

With the development of sophisticated interactive systems, computer interfaces increasingly have human-like appearances and are able to show emotional expressions or speak to their users. This has important consequences for the interactions that people have with those interfaces. Research has shown that people respond to computers as if they were social actors and even ascribe personalities to them (e.g., Nass et al., 1994; Nass, Moon, Fogg, Reeves, & Dryer, 1995; Reeves & Nass, 1996). Aristotle already argued that the arrangement of various parts of a speech (in this case social cues) is important in persuasive communication, but we still know relatively little about the effects of including multiple social cues in computer interfaces and how they influence people's responses (Surakka & Vanhala, 2011). This chapter explores the relation between consistency of social cues and people's susceptibility to social influence coming from non-human agents. I argue that people are more likely to attribute human-like characteristics to non-human agents that show consistent social cues than to those that show inconsistent social cues, and that they are more susceptible to social feedback coming from non-human agents that show consistent social cues.

3.1.1 Types of social cues

When humans interact with each other, they use a variety of social cues, often without being aware of them (Knapp, Hall, & Horgan, 2013). For example, they may use facial expressions and intonation of speech to communicate their emotions, and gestures and loudness of speech to emphasize certain aspects of the message (for an overview, see Knapp et al., 2013). In these social interactions, humans thus combine social cues that together make a message more understandable. The perceived meaning of combinations of social cues may be different from the sum of the individual cues. For example, when a person makes a sarcastic note, the intonation of his/her voice probably does not match his/her facial expression. In other words, if two social cues are perceived to be inconsistent with each other, this may influence people's interpretations of those cues, and consequently their persuasiveness. Understanding the combined effects of different social cues may help the design of social human-computer interactions.

Ever since Heider and Simmel (1944) showed that people attribute social intentions, characteristics and traits to moving geometric shapes, the importance of these different types of cues in social interactions has been thoroughly investigated. Examples of these types of social cues are motion dynamics, facial expressions, gazing behavior, and speech. Each of these cues could be used to communicate different kinds of social information. Motion dynamics provide cues for agency, as they could be used to infer an agent's mental states such as desires and intentions (e.g., C. D. Frith & Frith, 2007; U. Frith & Frith, 2010). Facial expressions help other people understand how a person feels (e.g., Adolphs, 2003), and these expressions are shown to be universal across cultures (Ekman, 1971). Gazing behavior could be used to learn what someone is thinking about (e.g., Langton, Watt, & Bruce, 2000; Emery, 2000; Allison, Puce, & McCarthy, 2000). Moreover, motivational orientations that belong to certain emotions can be inferred from people's gazing behavior (Argyle & Cook, 1976). In addition to non-verbal cues, verbal ones are being used as well. Speech is considered to be a crucial social cue in human-human interactions (Massaro, 1998), as well as in human-computer interactions (Nass & Gong, 2000).

3.1.2 Consistency of social cues

Research in social human-computer interactions has repeatedly investigated the effects of multiple social cues that were consistent with each other (e.g., Nass et al., 1994; Sproull et al., 1996; Blascovich et al., 2002; Vossen et al., 2010). For example, Vossen et al. (2010) studied the persuasive effects of speech and embodiment of a robot that provided feedback in an energy-saving task. In their experiment, the feedback was provided by a social cue (speech) or a non-social cue (colored light). As an additional cue, the feedback system used either social embodiment (a social robot) or non-social embodiment (a boxed computer). The feedback in the non-social embodiment condition was provided through speech files played by a computer. The feedback in the social embodiment condition used the same speech files, but also included facial expressions that matched the valence of the content of the speech files (Vossen et al., 2010). In other words, the persuasive effect of 'embodiment' was essentially a combined effect of providing two social cues that were consistent with each other: social embodiment and facial expressions.

I argue that social cues of artificial agents may not always be consistent with each other, for example when one or multiple cues are perceived as ambiguous. Furthermore, when only one of the social cues is manipulated (e.g. by providing positive speech with a neutral expression), this may be perceived as inconsistent. Also, when manipulating two different cues, one of them may accidentally be changed in the wrong direction, causing them to be perceived as inconsistent. These inconsistencies in social cues may decrease people's perceived human-likeness of artificial agents. Perceived inconsistencies in those cues may create confusion making people misunderstanding, misinterpreting, or incorrectly recalling aspects of the interaction. Consequently, such perceived inconsistencies could decrease the perceived human-likeness and the persuasiveness of artificial agents. For these reasons, investigating effects of consistency of social cues is important for the design of PT in the form of artificial agents.

One of the important functions of social interactions between humans is persuasion (e.g., Stiff & Mongeau, 2003). Research in social psychology has been designed to investigate how people influence each others' behavior. When creating a social interaction in which one of the humans is replaced by a computer (see Nass et al., 1994), the success of this interaction can be measured by the computer's persuasiveness. I argue that consistency of social cues determines people's attributions of human-like characteristics to artificial agents, which positively influences both people's recognition of those cues and their persuasiveness.

3.1.3 Research aims

The current research was designed with the aim to investigate effects of consistency of an artificial agent's social cues on people's recognition and recall of emotions conveyed by those agents and their persuasiveness. We hypothesized that an artificial agent that shows consistent social cues would make people perceive the agent as more human-like, to better recognize and recall its emotions and to be more persuaded compared to an artificial agent that uses inconsistent social cues. These hypotheses were investigated in two studies. Study 3 was designed to investigate effects of consistency of an artificial agent's social cues on the agent's perceived human-likeness and people's recognition of its emotions. To be able to compare effects of consistency of social cues for artificial agents and humans, the experimental setup of this study was adapted from earlier work on emotion recognition in human perception (Adams & Kleck, 2003, 2005). Study 4 was designed to investigate effects of consistency of an artificial agent's social cues on people's recall of the agent's social feedback and its persuasiveness.

3.2 Study 3

In this study, the effects of consistency of an artificial agent's social cues on the agent's perceived human-likeness and people's recognition of the agent's emotions were investigated. The two social cues that were used were gaze direction and facial expressions. Gaze direction is often related to the emotion that is experienced (e.g., Argyle & Cook, 1976; Kleinke, 1986). People show more *direct* gaze when they are seeking friendship or when they communicate threat, and they show more *averted* gaze as a result of heightened anxiety or increased depression (Kleinke, 1986). This indicates that a connection exists between the type of emotion experienced and people's gazing behavior. More specifically, gazing behavior is often used as a cue to express *approach*-oriented versus *avoidance*-oriented emotions (e.g. Argyle & Cook, 1976; Kleinke, 1986; Adams & Kleck, 2005).

In a series of studies, Adams and Kleck (2003, 2005) investigated the relation between gaze direction and facial expressions of emotion in human perception. They proposed that gaze direction as a social cue indicates a person's approach-avoidance behavioral tendencies (Adams & Kleck, 2005). That is, to form a consistent expression-gaze combination, approach-oriented emotions are most likely combined with a direct gaze, whereas avoidance-oriented emotions are most likely combined with an averted gaze. Therefore, direct gaze should increase the perception of approach-oriented emotions like anger and joy, whereas averted gaze should increase the perception of avoidance-oriented emotions like fear and sadness.

In the work by Adams and Kleck (2003), participants were instructed to indicate whether human faces displayed anger or fear (or, in a second study, joy or sadness) as quickly and accurately as possible. Results showed that, in the perception of human facial expressions, approach-oriented emotions (i.e. anger and joy) were more quickly recognized with a *direct* gaze, whereas avoidance-oriented emotions (i.e. fear and sadness) were more quickly recognized with an *averted* gaze (Adams & Kleck, 2003). Similar effects also occurred on trait attributions made to neutral faces and ambiguous facial blends (Adams & Kleck, 2005). More specifically, people attributed more anger and joy to neutral faces with a direct gaze, whereas they attributed more fear and sadness to neutral faces with an averted gaze (Adams & Kleck, 2005).

As a first step in the investigation of effects of consistency of social cues on people's perceptions of and responses to artificial agents, the current study was designed to conceptually replicate the findings by Adams and Kleck (2003). Following the paradigm presented by Nass et al. (1994), human faces were replaced with those of artificial agents. We hypothesized that a consistent expression-gaze combination of an artificial agent would be recognized more quickly and accurately than an inconsistent expression-gaze combination. Attributions of human-like characteristics to artificial agents were argued to be related to the perceived consistency of the agents' social cues. We therefore also hypothesized that artificial agents that show consistent expression-gaze combinations would be perceived as more human-like than artificial agents that show inconsistent expression-gaze combinations.

3.2.1 Method

Participants and design

Forty participants (25 males and 15 females; $M_{age} = 20.6$, $SD_{age} = 1.8$, Range = 18 to 26) were recruited. They received either course credit or €3 for their participation. The study consisted of three parts: An emotion recognition task, adapted from Adams and Kleck (2003) with a 2 (Expression: angry vs. sad) X 2 (Gaze-direction: direct vs. averted) within-subjects design, a 5-minute filler task that was unrelated to the current experiment, and a short questionnaire to measure anthropomorphism of the artificial agents used in the emotion recognition task.

Materials and procedure

At the start of the experiment, participants were informed via the computer screen that all collected data would be analyzed anonymously, and their rights to withdraw at any time (without consequences for payment) were explained. By pressing a key, they agreed to participate in the experiment.

Participants first performed the emotion recognition task. In this task, they were shown pictures of artificial agents, and had to indicate as quickly and correctly as possible whether the expressed emotion was either anger or sadness by pressing the 'A'- or 'L'-key. To control for effects of participants' dominant hand responses, labels of the categories were counterbalanced.

Pictures of four female and four male artificial agents were used. Examples of each of the artificial agents are shown in Table 3.1. Eight different pictures were generated for each of the artificial agents. Half of those pictures contained a sad expression and the other half an angry one. Also, half of them contained a direct gaze and the other half an

Table 3.1: Examples of each of the artificial agents used in the emotion recognition task in Study 3.

	Direct gaze		Averted gaze	
Angry expressions				
Sad expressions				

averted gaze. Each expression was displayed twice in both the averted gaze (left gaze and right gaze) and the direct gaze conditions to balance out the design, leading to a total of 64 different pictures. All pictures were displayed twice, making a total of 128 trials that were presented in random order. The dependent variables were response latencies and number of errors.

After participants performed the filler task, they rated the artificial agents' emotional expressions on their intensity, realism, and human-likeness on a scale ranging from 1 (not at all) to 7 (extremely) to measure anthropomorphism. Participants completed these three questions once for each of the eight artificial agents that were used in the emotion recognition task, leading to a total of 24 questions. The consistent (sad-averted and angry-direct) and inconsistent (sad-direct and angry-averted) expression-gaze combinations were equally distributed. The dependent variables were constructed by averaging the six responses on the angry-direct ($\alpha = .70$), angry-averted ($\alpha = .68$), sad-direct ($\alpha = .66$), and sad-averted ($\alpha = .81$) combinations.

This measuring instrument for anthropomorphism was different than those used in Chapter 2 for two reasons. First, the questions were going to be asked eight times, once for each of the artificial agents, so it would be more efficient to use a short questionnaire with clear and unambiguous concepts. Second, the questionnaires that were used in Chapter 2 were applied to interactive artificial agents, whereas still images were used in the current study. For this reason I refrained from including questions regarding concepts like the agents' free will and consciousness, which are unlikely to be attributed to artificial agents on still images. Furthermore, a question about for example the fluency of the agents' movements would not be applicable in the current setup.

At the end of the session, participants indicated their age and gender and left the room. Finally, they were debriefed, paid, and thanked for their contribution.

3.2.2 Results

Prior to further analyses, data on response latencies were log-transformed, which is a general approach for handling reaction time distributions (see e.g., Whelan, 2010). Trials resulting in incorrect responses (8.7%) were replaced by the mean response latency of all responses. Response latencies of one participant were slower than three standard deviations from the mean and data from this participant were excluded from further analyses. For ease of interpretation, response latencies were converted back into milliseconds for reporting the means and standard errors.

Next, effects of participants' dominant hand responses were checked by submitting the average log-transformed response latencies and the total number of errors to independent samples t-tests with the dominant hand responses (i.e. 'anger-dominant' vs. 'sad-dominant') as groups. Results showed no effects of dominance on either response latencies or number of errors, both t's < 1, both p's $> .32$. These findings indicated that participants' dominant hand responses did not influence their performance in the emotion recognition task.

Emotion recognition

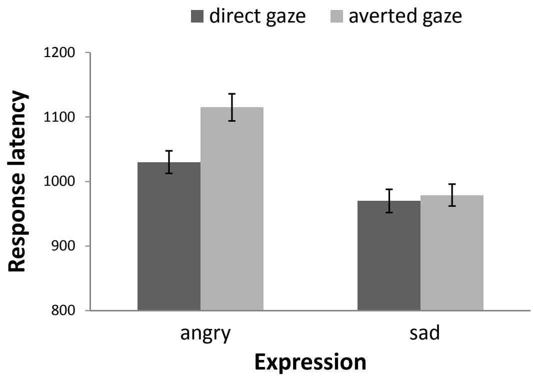
To test the hypothesis that consistent expression-gaze combinations would be recognized more quickly than inconsistent expression-gaze combinations, the log-transformed response latencies were submitted to a 2 (Expression: angry vs. sad) X 2 (Gaze-direction: direct vs. averted) analysis of variance (ANOVA). This analysis showed a significant main effect of Expression, $F(1, 38) = 24.17$, $p < .001$, ${\eta_p}^2 = .39$. More specifically, sad expressions ($M = 975$, $SE = 34.15$) were recognized more quickly than angry expressions ($M = 1072$, $SE = 37.06$). Furthermore, a significant main effect of Gaze-direction emerged, $F(1, 38) = 11.04$, $p < .01$, ${\eta_p}^2 = .23$. More specifically, expressions with a direct gaze ($M = 1000$, $SE = 32.84$) were recognized more quickly than expressions with an averted gaze ($M = 1047$, $SE = 36.81$). These main effects were qualified by the predicted interaction between Expression and Gaze-direction, $F(1, 38) = 7.18$, $p = .01$, ${\eta_p}^2 = .16$. More specifically, the consistent (sad-averted and anger-direct) combinations ($M = 1004$, $SE = 32.31$) were recognized more quickly than the inconsistent (sad-direct and anger-averted) ones ($M = 1042$, $SE = 36.95$). The averaged response latencies for each of the expression-gaze combinations are presented in Table 3.2 and visualized in Figure 3.1a.

Table 3.2: The averaged response latencies, number of errors and anthropomorphism ratings for each of the expression-gaze combinations in Study 3. Standard errors between brackets.

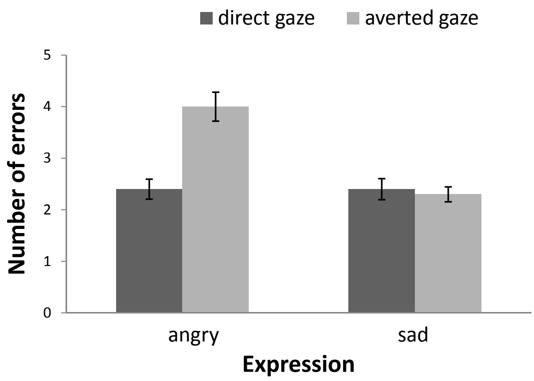
Expression	Response latency (in ms)		Number of Errors		Anthropomorphism	
	Direct	Averted	Direct	Averted	Direct	Averted
Angry	1030 (35)	1115 (42)	2.4 (0.39)	4.0 (0.56)	4.2 (0.12)	3.6 (0.14)
Sad	970 (36)	979 (34)	2.4 (0.41)	2.3 (0.29)	4.3 (0.13)	4.8 (0.15)

To test the hypothesis that consistent expression-gaze combinations would be recognized more accurately than inconsistent expression-gaze combinations, the total number of errors were submitted to a 2 (Expression:

angry vs. sad) X 2 (Gaze-direction: direct vs. averted) ANOVA. This analysis showed a significant main effect of Expression, $F(1, 38) = 4.37$, $p = .04$, $\eta_p{}^2 = .10$. More specifically, angry expressions ($M = 6.46$, $SE = 0.84$) were falsely recognized more often than sad expressions ($M = 4.67$, $SE = 0.65$). Furthermore, a significant main effect of Gaze-direction emerged, $F(1, 38) = 9.10$, $p < .01$, $\eta_p{}^2 = 0.19$. More specifically, expressions with an averted gaze ($M = 6.33$, $SE = 0.68$) were falsely recognized more often than expressions with a direct gaze ($M = 4.79$, $SE = 0.65$).



(a)



(b)

Figure 3.1: Visualization of the interaction between expression and gaze-direction for (a) the response latency, and (b) the number of errors in Study 3. Whiskers represent 95% error bars.

These main effects were qualified by the predicted interaction between Expression and Gaze-direction, $F(1, 38) = 7.17$, $p = .01$, ${\eta_p}^2 = .16$. More specifically, the inconsistent (sad-direct and anger-averted) combinations ($M = 6.38$, $SE = 0.80$) were falsely recognized more often than the consistent (sad-averted and anger-direct) ones ($M = 4.74$, $SE = 0.56$). The averaged number of errors for each of the expression-gaze combinations are presented in Table 3.2 and visualized in Figure 3.1b.

Anthropomorphism

To test the hypothesis that consistent expression-gaze combinations would be evaluated as more intense, realistic, and human-like than inconsistent expression-gaze combinations, the averaged levels of anthropomorphism were submitted to a 2 (Expression: angry vs. sad) X 2 (Gaze-direction: direct vs. averted) ANOVA. Results showed a significant main effect of Expression, $F(1, 38) = 26.36$, $p < .001$, ${\eta_p}^2 = .41$. More specifically, angry expressions ($M = 3.89$, $SD = 0.75$) were rated lower on anthropomorphism than sad expressions ($M = 4.51$, $SD = 0.77$). No significant main effect of Gaze-direction was found, $F(1, 38) = 0.80$, $p = .38$.

Figure 3.2: Visualization of the interaction between expression and gaze-direction for anthropomorphism ratings in Study 3. Whiskers represent 95% error bars.

Finally, a significant interaction between Expression and Gaze-direction emerged, $F(1, 38) = 47.98$, $p < .001$, $\eta_p{}^2 = .56$. More specifically, the consistent (anger-direct and sad-averted) combinations ($M = 4.52$, $SD = 0.68$) were rated higher on anthropomorphism than the inconsistent (anger-averted and sad-direct) ones ($M = 3.88$, $SD = 0.76$). The averaged anthropomorphism levels for each of the expression-gaze combinations are presented in Table 3.2 and visualized in Figure 3.2.

3.2.3 Discussion

This study was designed to investigate effects of consistency of an artificial agent's social cues on people's recognition of the agent's emotions and their anthropomorphism. We hypothesized that consistent expression-gaze combinations would be more quickly and accurately recognized than inconsistent expression-gaze combinations. More specifically, angry expressions were expected to be recognized more quickly and accurately when combined with a direct gaze, and, in contrast, sad expressions were expected to be recognized more quickly and accurately when combined with an averted gaze. In addition, we hypothesized that people would anthropomorphize consistent expression-gaze combinations to a greater extent than inconsistent expression-gaze combinations. The results supported these expectations, conceptually replicating earlier findings on facial expressions of emotion in human perception (Adams & Kleck, 2005, 2003). Participants more quickly and accurately recognized consistent expression-gaze combinations than inconsistent ones. Additionally, consistent expression-gaze combinations were perceived with higher levels of anthropomorphism than inconsistent ones.

In the current study, participants' response latencies were on average much slower than those found by Adams and Kleck (2003). One notable difference between the studies is that Adams and Kleck (2003) used a fixation point that marked the position of the presented stimulus to increase the participants' focus on the center of the screen, whereas in the current study such a fixation point was not included. Nevertheless, findings of the current study replicated those of Adams and Kleck (2003), indicating that slower response latencies did not influence the consistency effect.

Interestingly, effects of consistency on response latencies and accuracy were found to be present mainly for the angry expressions, but not for the sad ones. This could have occurred due to a ceiling effect, because the sad expressions were recognized much more quickly than angry ones. I argue that sad expressions with a *direct* gaze (i.e. showing inconsistent cues) were already recognized so quickly that changing the gaze direction to be consistent with the expression could not increase participants' recognition speed. It would be interesting to investigate the extent to which response latencies for different types of emotional expressions differ, but this is outside of the scope of the current chapter.

The consistency effect was found for combinations of gaze direction and emotional expressions, but it may be generalized to other types of social cues as well. For example, speech is considered to be an important determinant for people's responses in social human-computer interactions (Nass & Gong, 2000). Moreover, when delivering an interactive persuasive message, speech may be a more suitable social cue than gazing behavior. It may therefore be valuable to include speech as a social cue in the design of artificial agents that are aimed at influencing people's behavior.

In the current study, anthropomorphism was measured by the extent to which the artificial agents' expressions were perceived as intense, realistic, and human-like. Results showed that consistency of social cues positively influenced these concepts. Because of their higher perceived human-likeness, artificial agents that show social cues that are consistent with each other may also be more effective in influencing people's behavior in an interactive setting. Study 4 was designed with the aim to extend the findings of the current study to a different type of social cue (i.e. speech), and to investigate whether consistency of social cues could also make an artificial agent more persuasive. Because of the possible ceiling effect of using sad expressions in the current study, the second study refrained from using anger and sadness as the two emotions and used anger and happiness instead.

3.3 Study 4

In this study, the effects of consistency of an artificial agent's social cues on people's recall of the agent's social feedback and its persuasiveness were investigated. The two social cues used were speech and facial expressions. The use of speech as a social cue has been linked to socio-evolutionary principles, because it is argued to be the most prevalent cue of humanness (e.g., Nass & Gong, 2000; Nass & Brave, 2005). Using speech in human-computer interactions could also strongly influence people's experiences of those interactions. For example, artificial agents using speech elicited stronger feelings of social presence than text-based interactions (Qiu & Benbasat, 2009). That is, using a human voice significantly increased people's feelings of social presence, compared with using written text only or text-to-speech (Qiu & Benbasat, 2009). This finding was in line with earlier findings on audio- or video-conference interactions, which were shown to elicit stronger feelings of perceived social presence than text-chat interactions (Sallnäs, 2005).

The current study was designed to extend the findings of Study 3. We hypothesized that consistency of expression-speech combinations of artificial agents would increase participants' recall of the agent's social feedback and its persuasiveness. Results of Studies 1 and 2 showed that females were more susceptible to persuasion by artificial agents than males. Therefore, effects of participant gender were explored in the current study as well.

3.3.1 Method

Participants and design

Seventy people (38 males and 32 females; $M_{age} = 21.9$, $SD_{age} = 4.5$, Range = 17 to 46) were randomly assigned to one of two experimental conditions in a between-subjects design with consistent ($n = 35$, 19 males and 16 females) and inconsistent ($n = 35$, 19 males and 16 females) expression-speech combinations as groups. The experiment lasted for 30 minutes for which participants were paid €5.

Materials and procedure

At the start of the experiment, participants were informed via the computer screen that all collected data would be analyzed anonymously, and their rights to withdraw at any time (without consequences for payment) were explained. By signing a form, they agreed to participate in the experiment.

Figure 3.3: Screenshot of the simulated thermostat interface (left) and artificial agent Kim (right) as used in Study 4.

Participants were first introduced to artificial agent Kim (see Figure 3.3) that would provide feedback about the choices they made during the thermostat task as developed by Ham, Midden, Maan, and Merkus (2009). In this task, a simulated thermostat interface (see Figure 3.3) was presented on the screen and participants were asked to complete ten heating tasks (see Appendix C for a list of these tasks).

During the setting of the simulated thermostat, participants received feedback from the artificial agent, using both speech and facial expressions. The speech feedback was based on the energy that was used by the chosen settings and had six levels, three of which positive (e.g., 'Your setting of the thermostat is fantastic!') and three negative (e.g., 'Your total energy use is terrible!'). The expressions had two levels, one positive (happy) and one negative (angry). The artificial agent used direct gaze in both conditions to match with the approach-oriented emotions happiness and anger.

In the consistent condition, the expressions matched the spoken feedback. More specifically, positive speech was combined with a positive (happy) expression. In the inconsistent condition, the expressions did not match the spoken feedback. Thus, positive speech was combined with a negative (angry) expression. Participants completed two practice trials and ten experimental trials that were presented in random order, and received social feedback from the artificial agent every time they completed a heating task by pressing a 'Finished'-button.

The main dependent variable was created by standardizing the energy used with the chosen settings. This value shows the immediate effect of the agent's feedback on participants' energy use. For ease of interpretation, energy use was converted to positive numbers to represent Energy-conservation.

For measuring anthropomorphism, a tentative version of the anthropomorphism scale that was developed in Chapter 4 was included. This version of the scale consisted of 33 statements about human-like characteristics that could be attributed to the artificial agent (e.g., 'Kim deliberately performs actions.'). The statements could be answered with 'no' (coded as a 0) and 'yes' (coded as a 1), and participants' responses were averaged into one anthropomorphism score. No effects were found on this scale ($F(1, 69) = 0.40$, $p = .53$) and it will not be included in the results section. More details and discussion about the development, benefits, and limitations of this scale are presented in Chapter 4.

To measure participants' recall of the feedback that was provided by the artificial agent, they were asked three questions. These questions were about the general content of the feedback, the agent's emotional expression and the agent's speech feedback, all on 6-point scales. To prevent predictability, these questions were asked after the third, sixth, and tenth trial. The absolute difference between participants' responses and the correct one was calculated and two scores were created, one for the general content of the feedback and one for the agent's emotional expression[4]. For ease of interpretation, the scores were coded such that higher values represent more accurate recall.

At the end of the session, participants indicated their age, gender, and other demographics including the size of their household and their occupation. After this, they were debriefed, paid and thanked for their contribution.

3.3.2 Results

Feedback recall

To test the hypothesis that consistent expression-speech combinations would make participants better recall the feedback than inconsistent expression-speech combinations, the two recall scores were submitted to independent-samples t-tests with the two Consistency conditions as

[4]Due to a program error, data on the speech feedback question were not saved properly. Therefore, only data on the questions about the general content of the feedback and the agent's emotional expression could be included in the analysis.

groups. Results showed a significant difference between the groups on the general recall of the feedback, $t(68) = 2.17$, $p = .03$, $r^2 = .06$. More specifically, participants in the consistent condition ($M = 2.85$, $SD = 0.76$) more accurately recalled the general feedback than participants in the inconsistent condition ($M = 2.42$, $SD = 0.88$). This effect is visualized in Figure 3.4a.

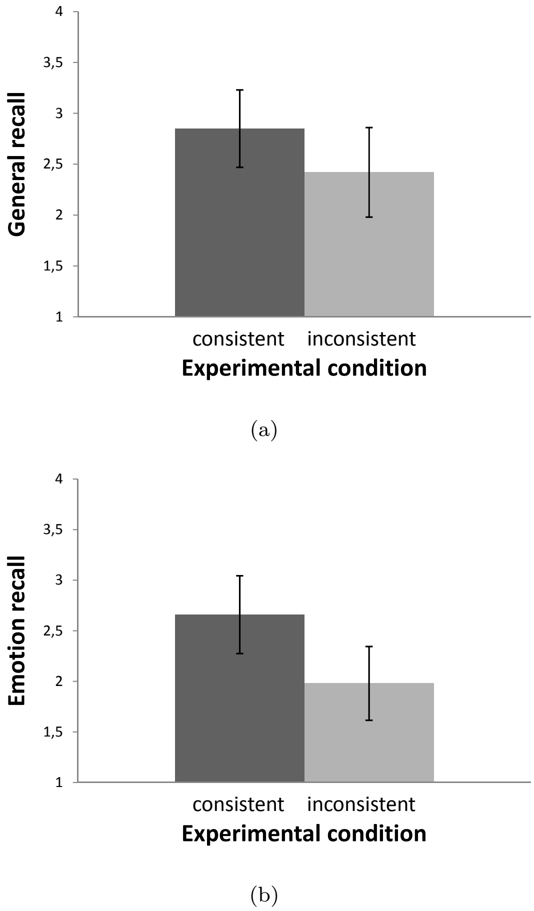


Figure 3.4: Visualization of participants' recall accuracy on (a) the general feedback and (b) the agent's emotional expressions per experimental condition in Study 4. Whiskers represent 95% confidence intervals.

Results also showed a significant difference between the groups on recall of the agent's emotional expressions, $t(68) = 4.07$, $p < .001$, $r^2 = .20$.

More specifically, participants in the consistent condition (M = 2.66, SD = 0.77) more accurately recalled the agent's emotional expressions than participants in the inconsistent condition (M = 1.80, SD = 0.98). This effect is visualized in Figure 3.4b.

Energy conservation

The hypothesis that consistent expression-speech combinations would increase participants' susceptibility to persuasion from an artificial agent was tested with a Linear Mixed Model (LMM) with a single factor (Consistency: consistent vs. inconsistent) between-subjects design with Energy-conservation as dependent variable and the specific heating task as random factor. This analysis showed a marginally significant effect of Consistency, $F(1, 638.204) = 2.51$, $p(one\text{-}sided) = .06$, $\omega^2 = .002$. More specifically, participants who received feedback with consistent expression-speech combinations ($EMM = 0.38$, $SE = 0.07$) saved more energy than participants who received feedback with inconsistent expres-sion-speech combinations ($EMM = 0.28$, $SE = 0.07$). This effect is visualized in Figure 3.5.

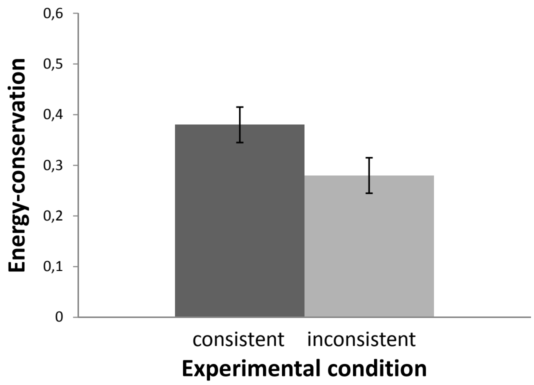


Figure 3.5: Visualization of energy-conservation per experimental condition in Study 4. Whiskers represent 95% confidence intervals.

To investigate the effect of Participant gender on susceptibility to social influence, Energy-conservation was submitted to a 2 (Consistency: consistent vs. inconsistent) X 2 (Participant gender: male vs. female) LMM with the specific heating task as random factor. Results showed a significant main effect of Participant gender, $F(1, 640.09) = 31.32$, $p < .001$, $\omega^2 = .005$. More specifically, female participants ($EMM = 0.51$, $SE = 0.07$) saved more energy than male participants ($EMM = 0.13$, $SE = 0.07$). This effect is visualized in Figure 3.6. Results provided no evidence for an interaction between Consistency and Participant gender, $F(1, 636.50) = 0.09$, $p = .77$.

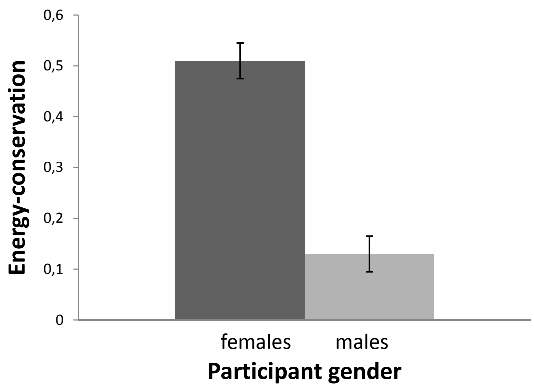


Figure 3.6: Visualization of energy-conservation per participant gender in Study 4. Whiskers represent 95% confidence intervals.

3.3.3 Discussion

This study was designed to investigate effects of consistency of an artificial agent's social cues on people's recall of the agent's social feedback and its persuasiveness. We hypothesized that consistent expression-speech combinations would make people better recall feedback provided by an artificial agent compared with inconsistent expression-speech combinations. In addition, we hypothesized that people would adapt their behavior more as a result of consistent expression-speech feedback than inconsistent expression-speech feedback. Results showed that participants better recalled the feedback when it was provided using consistent expression-speech combinations. These results extended the consistency effect found in Study 3 to a different type of social cue. Moreover, participants who received consistent expression-speech feedback showed more behavior change as a result of the feedback than those who received inconsistent expression-speech feedback. Finally, females appeared to be more susceptible to the artificial agent's persuasion than males.

In line with findings of Study 3, consistent expression-speech combinations increased participants' recall of feedback provided by the artificial agent compared with inconsistent expression-speech combinations. This finding shows that the effect of consistency on recognition is not limited to expression-gaze combinations, but also occurs for expression-speech combinations. Additionally, this finding shows that the consistency effect occurs not only when people were asked to quickly recognize an artificial agents' emotions, but also when they had to recall (aspects of) an interaction afterwards. Results did not support the hypothesis that artificial agents that show consistent expression-speech combinations would be perceived as more human-like than those that show inconsistent expression-speech combinations. The anthropomorphism scale was still in development at this point. More details on the development of this scale will be presented in Chapter 4.

In addition to the effects of consistency of an artificial agent's social cues on people's recall of its feedback, consistency also influenced participants' energy use in the thermostat task. This finding shows that the consistency effect was not limited to recognition of artificial agents' emotional expressions, but also influenced that agent's persuasive effectiveness. This is an important finding for the design of PT in the form of artificial agents, because it shows the importance of consistency of the agent's social cues for its persuasiveness.

The effect of consistency on participants' behavior was however limited, possibly because participants mainly focused on one of the two social cues. In contrast to the facial expressions, the speech cue always correctly represented participants' energy use, and this cue was difficult to ignore. If participants indeed predominantly focused on the speech cue, this could have influenced the strength of the consistency effect. Future research could be designed to expand the current design with an inconsistent condition in which the emotional expressions always correctly represent participants' energy use and the speech is either correct or incorrect.

3.4 General discussion

The current research was designed to investigate effects of consistency of social cues as a technological determinant of anthropomorphism on people's recognition and recall of emotions conveyed by artificial agents and their persuasiveness. Earlier research on social cues only investigated cues that were consistent with each other. When social cues are perceived as inconsistent, this could cause confusion and make people misunderstanding, misinterpreting, or incorrectly recalling aspects of the interaction. The effects of consistency of social cues were explored in two studies. Results of Study 3 showed an effect of consistency of social cues on people's speed and accuracy of recognizing emotional expressions. Results of Study 4 showed effects of consistency of social cues on people's ability to recall aspects of persuasive messages, and their susceptibility to social feedback in an energy-saving task.

Study 3 investigated whether consistency of social cues influenced people's performance in recognizing artificial agents' emotions. Results showed that consistency of two social cues (i.e. facial expressions and gaze direction) increased people's recognition speed and accuracy of artificial agents' emotional expressions, and their anthropomorphism. These results conceptually replicated earlier findings by Adams and Kleck (2003) and extended the consistency effect found in human-human interactions to the domain of human-agent interactions. Additionally, anthropomorphism of the artificial agents was found to be higher for those that showed consistent social cues than those that showed inconsistent ones. This finding indicated that inconsistencies of social cues may influence people's perceptions of artificial agents, which could also make them less persuasive.

Study 4 investigated whether consistency of social cues indeed influenced people's susceptibility to persuasion by artificial agents. Results showed that that consistency of two social cues (i.e. facial expressions and speech) increased recall of social feedback provided by an artificial agent and it influenced people's behavior in an energy-saving task. These findings extended the consistency effect found in Study 3 to a different type of social cue (i.e. speech), and to an interactive setting.

3.4.1 Limitations and future research

In both studies, multiple social cues were used that were either consistent or inconsistent with each other. However, the cues that were used differed between the two studies. Facial expressions were used in both studies, but the first study used gazing, whereas the second study used speech as a second cue. On the one hand, the replication of the consistency effect could be interpreted as the effect being generalizable to different types of social cues. However, the effects found in both studies could also be caused by different underlying processes. Different areas of the human brain are found to be responsible for recognition of facial expressions than for responses to persuasive messages (for an overview, see e.g., Roland, 1993). Expanding the design of the current studies to include more types of social cues could provide insight into how these different processes influence the consistency effect.

In addition to the differences in types of social cues that were used in the two studies, participants also performed different tasks. In Study 3 they performed an emotion recognition task, whereas in Study 4 they performed an interactive energy-saving task. Quick and accurate recognition and categorization of emotional expressions may require different processing than deliberately changing the temperature settings on a simulated thermostat interface. Replication of the consistency effect could be interpreted as it being generalizable to different types of tasks, but more work is needed to investigate circumstances in which the effects of consistency of social cues of artificial agents could break down. For example, when people perform very easy tasks or when one of the social cues becomes dominant, the consistency effect may decrease or even completely disappear. For the design of PT in the form of artificial agents, it is important to understand when and how consistency of social cues is most likely to influence the user's responses to those agents.

3.4.2 Conclusions

Overall, results from the current chapter are promising and adding to the body of literature in the domain of PT in the form of artificial agents. Results showed that an agent's persuasiveness can be influenced by the combined meaning of the social cues it shows, which suggests that developing artificial agents that have multiple social cues which are consistent with each other could contribute to the persuasiveness of those agents. Reversely, when people misinterpret one or more of the social cues, their effectiveness could be seriously hampered. Adding social cues to a computer interface could improve its effectiveness, as long as the (perceived) meanings of the individual cues match.

Chapters 2 and 3 were designed to investigate the relation between determinants of anthropomorphism and an artificial agent's persuasiveness. Findings of Chapter 2 indicated that the psychological state of a user could influence a technology's persuasive effectiveness. Findings of the current chapter indicated that the combined meaning of social cues could also influence a technology's persuasive effectiveness. These findings pose design challenges for PT, in particular those in the social actor role, because they are designed to take part in social interactions with their users.

In the studies in Chapters 2 and 3, both *determinants* and *outcomes* of anthropomorphism were included. Findings confirmed the expected relation between those determinants and persuasion. However, results did not demonstrate the role of anthropomorphism itself. I argue that the measuring instruments for anthropomorphism that were used in these two chapters may not be sufficient. The next chapter will address this issue of measuring anthropomorphism.

"What a familiarity with the construction of Turing test bots had begun to show me was that we fail –again and again– to actually be human with other humans, so maddeningly much of the time."

Brian Christian, The Most Human Human (2011)

4

Measuring Perceived Human-likeness

The development of the Predisposition to Anthropomorphize Scale

This chapter is largely based on:

Ruijten, P. A. M., Haans, A., Ham, J., and Midden, C. J. H (2015). *Measuring Perceived Human-likeness: The development of the Predisposition to Anthropomorphize Scale.* Manuscript in preparation.

Ruijten, P. A. M, Bouten, D. H. L., Rouschop, D. C. J., Ham, J., & Midden, C. J. H. (2014). Introducing a rasch-type anthropomorphism scale. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction* (pp. 280-281). ACM New York 2014.

This chapter explores whether reliable measurement of anthropomorphism can be improved. Based on findings of the previous two chapters, I argue that the available measuring instruments for anthropomorphism may be insufficient. This argument is based on the notion that different researchers appear to focus on different subsets of human-like characteristics, because they interpret the concept in different ways. In the current chapter, I discuss these different interpretations and develop a new measuring instrument based on the Rasch model (e.g., Bond & Fox, 2013) that aims to measure anthropomorphism as a wide range of human-like characteristics on a one-dimensional scale. To test its validity, this new measuring instrument will be compared with two available ones. To test its sensitivity for differentiating between agent types, people's responses to four different kinds of agents will be compared.

4.1 General introduction

With technology playing an increasingly prominent role in our lives, our interactions with it become more important as well. In the near future, technology may get more human-like features and is likely to be represented as social entities that have faces, can adapt to our moods, and may even engage in conversation with us. In his book The Most Human Human, Brian Christian describes his experiences of being a confederate in the yearly Turing test, in which sophisticated software programs try to convince judges that they can 'think' (Christian, 2012). The computer that fools most judges is the most human-like one and wins the prestigious Most Human Computer award.

Thus, technology is becoming become more and more like us, as people perceive them as more human-like. The term anthropomorphism refers to the degree to which people perceive non-human objects (e.g., technology, robots) as human-like (e.g., Kennedy, 1992; Duffy, 2003; Epley et al., 2007; Bartneck et al., 2009; Waytz, Morewedge, et al., 2010; Eyssel, Kuchenbrandt, & Bobinger, 2011). This conventional and rather general description of anthropomorphism has lead to a variety of interpretations of the concept by different researchers, thereby often referring to different subsets of human-like characteristics.

4.1.1 Subsets of human-like characteristics

Human-like characteristics include appearances, thoughts and emotions. *Appearances* are characteristics that reflect human form or behavior (i.e. how the object or robot looks and/or moves), including both physical shapes and physical abilities. With their measuring instrument for anthropomorphism, Bartneck et al. (2009) focused on such appearances by asking people to indicate to what extent a robot looks human-like, looks life-like, and shows realistic movements (Bartneck et al., 2009).

Thoughts are characteristics that reflect cognitive states and processes. According to Waytz, Morewedge, et al. (2010), anthropomorphism is a process of inductive inference which most likely occurs by attributing cognitive states that are perceived to be uniquely human to other agents (for a review, see Epley et al., 2007). Waytz, Morewedge, et al. (2010) measured anthropomorphism by asking people to indicate to what extent an agent has for example consciousness and free will.

Emotions are characteristics that indicate subjective conscious experiences and can be distinguished in primary and secondary ones (for an overview of the hierarchical organization of emotions, see Shaver, Schwartz, Kirson, & O'Connor, 1987). Eyssel, Hegel, Horstmann, and Wagner (2010) measured anthropomorphism by asking people to indicate to what extent robots can experience such primary and secondary emotions.

In their further work, Eyssel and colleagues differentiated between personality traits that reflect human nature and human uniqueness (Eyssel et al., 2011; Eyssel & Reich, 2013). This distinction between human nature and human uniqueness was adapted from earlier research on social perception in humans (Haslam et al., 2005, 2008). In this research, human nature characteristics are described as characteristics of the human species that are shared with other animals (e.g., innate and affective traits). Uniquely human characteristics, on the other hand, are considered to be exclusive to humans and not possessed by any other species (e.g., social learning and higher cognition, Haslam et al., 2008). This distinction between human nature and human uniqueness traits has lead to two-dimensional approaches of anthropomorphism.

4.1.2 Two-dimensional approaches

Based on earlier research on social perception in humans (Haslam et al., 2005), one could approach anthropomorphism as a two-dimensional construct. For example, Eyssel et al. (2011) measured anthropomorphism with 20 personality traits that were divided into two 10-item scales that reflect human nature and human uniqueness (adapted from Haslam et al., 2008). They argued that individuals who lack human uniqueness traits are implicitly perceived as animals, and that those who lack human nature traits are implicitly perceived as machines. Eyssel et al. (2011) used those two sets of personality traits for measuring anthropomorphism, but they did not consider them as separate dimensions of the construct. Findings on both scales were highly similar (Eyssel et al., 2011). In fact, Eyssel and Reich (2013) combined those two scales into one 'human essence' scale ($\alpha = .86$), suggesting that human nature and human uniqueness are not independent of each other, but rather represent certain levels of anthropomorphism.

In contrast, Złotowski, Strasser, and Bartneck (2014) argued that human nature and human uniqueness traits should be distinguished to better understand the impact of anthropomorphism on human-robot interactions. That is, they approached anthropomorphism as a two-dimensional construct with human nature and human uniqueness as two separate dimensions (Złotowski et al., 2014). However, when they tested this dimensionality of anthropomorphism, a two-factor model showed a bad fit of the data (Złotowski et al., 2014). Results also showed that human nature and human uniqueness were significantly correlated overall ($r = .55$, $p < .001$), but not for subjects who rated a robot high on either human nature or human uniqueness. For this reason, the authors suggested that two separate dimensions of anthropomorphism exist (Złotowski et al., 2014).

Finally, Waytz, Cacioppo, and Epley (2010) developed a scale to measure stable individual differences in anthropomorphism. They generated a list of items in which each of four types of agents (i.e. nonhuman animals, natural entities, spiritual agents such as gods and ghosts, and technological devices) were combined with five anthropomorphic traits (i.e. having intentions, free will, consciousness, a mind of its own, and experiencing emotions) to create a 20-item scale. When the questions about spiritual agents were excluded from this scale, a factor analysis revealed a two-factor solution as optimal, with a distinction between anthropomorphism of non-animal versus animal stimuli. However, these two factors were positively correlated ($r = .52$, $p < .001$), and further analysis revealed that a model with 'general anthropomorphism' as the only factor provided a good fit (Waytz, Cacioppo, & Epley, 2010). The authors concluded that anthropomorphism of animal and non-animal stimuli appear to be instances of a more general tendency to anthropomorphize (Waytz, Cacioppo, & Epley, 2010). Moreover, a possible two-factor solution for the scale indicates different categories of *agent types* rather than human-like characteristics.

Taken together, findings on the dimensionality of anthropomorphism do not provide convincing evidence for a two-dimensional structure of the concept. Human nature and human uniqueness may rather be viewed as different levels of anthropomorphism. Researchers who approached anthropomorphism as a one-dimensional construct seemed to focus on different *subsets* of human-like characteristics in their measures. In particular, Bartneck et al. (2009)'s measuring instrument for anthropomorphism mainly consists of appearance-related characteristics whereas Waytz, Morewedge, et al. (2010)'s measuring instrument mainly consists of thought-related characteristics. Since these instruments appear to only measure a subset of human-like characteristics –each focusing on a limited range–, measurements obtained with these instruments may have little correspondence. As a result, the question arises to what extent these different scales measure the same concept, and thus ultimately whether we can faithfully compare research findings. This inability to compare research findings on anthropomorphism seriously hampers the understanding of the concept and the design of social interactions between humans and their technology.

In this chapter, I aim to contribute to a solution of this problem. I will propose a one-dimensional conceptualization of anthropomorphism, and develop a novel measuring instrument (i.e. the Predisposition to Anthropomorphize Scale) based on the Rasch model (see e.g., Bond & Fox, 2013). Next, this instrument's validity will be tested by comparing it to the concepts of human nature and human uniqueness, and to existing measuring instruments of anthropomorphism. Finally, the instrument's sensitivity will be tested by comparing its ability to differentiate between different kinds of agents.

4.1.3 Conceptualizing anthropomorphism

I argue that the concept of anthropomorphism has three specific properties. First, anthropomorphism is a one-dimensional construct, and all human-like characteristics –no matter which subset they belong to– are ordered according to *the probability with which they are ascribed* to robots. Some human-like characteristics will be more easily ascribed to robots than others. For example, appearances will be more easily ascribed to robots than underlying cognitive states and processes, regardless of an individual's general tendency to anthropomorphize.

Second, I argue that this ordering of human-like characteristics is *similar for all individuals*. All people are expected to attribute more human-like appearances than cognitive states to robots. Such an invariant ordering also entails that if a person attributes the ability of moral reasoning to a robot, he/she is also expected to attribute the ability of seeing to that robot. Another person who does not attribute the ability of seeing to the same robot is not expected to attribute the ability of moral reasoning to it.

Third, I argue that anthropomorphism is *a single human predisposition*. More specifically, whatever human-like characteristics an individual ascribes to a robot –be it human nature or human uniqueness characteristics, basic physical abilities or moral decision making–, they all stem from that single predisposition to do so. This single predisposition was described as the tendency of attributing human-like characteristics in Chapter 1.

If all human-like characteristics can be invariantly ordered across people, those people can be compared on their individual predisposition to anthropomorphize on a single dimension, and simultaneously the anthropomorphism of different robots can be compared as well. One model that is able to map a person's predisposition to anthropomorphize and the human-like characteristics he/she is likely to attribute to a robot as locations on a single dimension, and thus seems highly suitable for measuring anthropomorphism, is the Rasch model (e.g., Bond & Fox, 2013; Haans, Kaiser, Bouwhuis, & IJsselsteijn, 2012).

4.1.4 The Rasch model

The Rasch model (see Equation 4.1) describes the natural logarithm of the odds of attributing a specific human-like characteristic i as an additive function of a person n's general predisposition to anthropomorphize (θ_{n}) and the difficulty to attribute that specific human-like characteristic to a robot (δ_{i}).

$$\ln\left(\frac{P(x_{ni}=1)}{1-P(x_{ni}=1)}\right)=\theta_n-\delta_i \tag{4.1}$$

The model has two basic assumptions. First, a person with a higher predisposition to anthropomorphize is expected to have a higher chance of attributing any human-like characteristic to a robot than a person with a lower predisposition to anthropomorphize. Second, all people are expected to have a higher chance of attributing an item low in human-likeness to a robot than of attributing an item high in human-likeness to that same robot. These assumptions match with the properties of anthropomorphism.

Both parameters in the equation (predisposition to anthropomorphize θ, and difficulty δ) are expressed on an interval scale in log odd units (also called logits, see Bond & Fox, 2013). For a specific human-like characteristic i to have a 50% chance to be attributed to a robot, its difficulty (e.g., $\delta_{\mathrm{i}}=1$) has to be matched numerically with an equivalent amount of a person n's predisposition to anthropomorphize ($\theta_{\mathrm{n}}=1$). For more detailed explanations of the Rasch model, see e.g., Bond and Fox (2013) and Haans et al. (2012).

The greatest advantage of using the Rasch model for measuring anthropomorphism is its potential to simultaneously compare the anthropomorphism of different robots, and people's predispositions to anthropomorphize on a single dimension. Moreover, the Rasch model assumes unidimensionality in the data, and reports the extent to which the data match this expected unidimensionality. An additional advantage of the model is that by using a so-called 'item bank', items in the measuring instrument can be deleted and/or replaced without influencing the instruments' validity, which enables the use of different sets of items in different studies (see Wright, 1977).

4.1.5 Research aims

The current research was designed with the aim to test the hypotheses that anthropomorphism can be successfully mapped onto a one-dimensional scale and that human-like characteristics are ordered in a way that is similar for all individuals in their encounter with different types of agents in different contexts. These hypotheses were investigated in three studies.

In Study 5, a 37-item Predisposition to Anthropomorphize Scale was developed and tested on its construct validity. Construct validity refers to the relation between people's responses to the items on the Predisposition to Anthropomorphize Scale and their perceived human nature and human uniqueness. Construct validity is high when strong relations between the items on the scale and these concepts are found. Because of the expected unidimensionality, human nature and human uniqueness were also expected to be strongly correlated. Additionally, items were expected to cover a wide range on the human-likeness continuum (which ranges from low to high in human-likeness) while being ordered according to the probability of being ascribed to robots.

In Study 6, a 25-item version of the Predisposition to Anthropomorphize Scale was tested on its convergent validity. Convergent validity refers to the extent to which estimates of the Predisposition to Anthropomorphize Scale are related to estimates obtained with two available measuring instruments. The instruments that were used are the questionnaire developed by Waytz, Morewedge, et al. (2010, the Waytz-instrument, see Appendix A.2) and the anthropomorphism part of the Godspeed questionnaire developed by Bartneck et al. (2009, the God-speed-instrument, see Appendix A.1). These two instruments appear to cover different subsets of human-like characteristics, each focusing on a limited range. The Predisposition to Anthropomorphize Scale was expected to moderately correlate with the Waytz- and Godspeed-instru-ments, and to measure anthropomorphism on a wider range of the human-likeness continuum than the two available instruments. In addition, two different robots were evaluated on their perceived human-likeness, and an invariant ordering of the items on the Predisposition to Anthropomorphize Scale for those two robots was expected.

In Study 7, a 19-item version of the Predisposition to Anthropomorphize Scale was tested on its sensitivity for differentiating between agent types. Sensitivity refers to the scale's ability to differentiate between anthropomorphism of different kinds of agents: humans, robots, computers, and algorithms. The scale's sensitivity was compared to that of the Waytz-instrument. To create a social interaction with the different agents, participants played the Ultimatum Game (e.g., Güth, Schmittberger, & Schwarze, 1982). The Predisposition to Anthropomorphize Scale was expected to successfully differentiate between all four different types of agents.

Finally, the Predisposition to Anthropomorphize Scale was expected to show an invariant ordering of human-like characteristics on a single dimension across all conditions and all studies.

4.2 Study 5

In this study, a first version of the Predisposition to Anthropomorphize Scale was developed and tested on its construct validity. A list of 37 human-like characteristics was created, largely based on work on humanness and anthropomorphism (e.g., Haslam et al., 2005, 2008; Bartneck et al., 2009; Waytz, Morewedge, et al., 2010; Eyssel et al., 2010). These characteristics were modeled as a function of a person's predisposition to anthropomorphize and the difficulty to attribute that human-like characteristic to a robot. We hypothesized that items and persons can be mapped onto a single one-dimensional scale, and that the items would be invariantly ordered according to the difficulty with which they were attributed to a robot.

Additionally, for construct validity purposes, the extent to which the 37 human-like characteristics were perceived as being human nature and uniquely human was measured. We hypothesized that the estimated difficulties with which the 37 characteristics are attributed to a robot would be related to their perceived human nature and human uniqueness. Because of the expected unidimensionality, human nature and human uniqueness were expected to be strongly correlated as well.

4.2.1 Method

Participants and design

One hundred and sixty one participants sampled through social media participated in one of three groups in the current study. The first group consisted of 124 participants (53 males and 71 females; M_{age} = 26.08, SD_{age} = 8.82, Range = 15 to 59) who were given a description about a robot and completed 37 survey items. Another group of 20 participants (9 males and 9 females, M_{age} = 19.94, SD_{age} = 1.98, Range = 18 to 23; two participants did not indicate their age and gender) rated the 37 human-like characteristics on human nature. The remaining 17 participants (11 males and 6 females, M_{age} = 21.12, SD_{age} = 1.80, Range = 18 to 24) rated all characteristics on human uniqueness. Participants in all three groups participated voluntarily, gave informed consent, and were not compensated for participation.

Materials and procedure

A set of 37 items describing human-like characteristics was constructed (the Predisposition to Anthropomorphize Scale, see Appendix A.8). These characteristics were aimed to represent abilities with various levels of humanness, ranging from low (e.g. detecting objects or estimating distances) to middle (e.g. recognizing voices or being self-conscious) to high (e.g. empathizing or having free will). For all three groups of participants, items were arranged in alphabetic order.

For the first group of 124 participants, items were formulated as a statement which could be answered with yes (coded with a 1) or no (coded with a 0). The items were presented through an online survey. The first page of the survey contained a short explanation about the study. On the next page, a short description about a robot was given: 'The robot has eyes to perceive the environment, has arms and legs to move around in this environment, and today the robot is trying to solve a moral dilemma'. This description was followed by an instruction to not think elaborately about the statements and to give the answers that first jumped to mind. Finally, participants indicated their gender, age and education level, and they were thanked for their contribution. This study took approximately 5 minutes to complete.

The two other groups of participants were not given the description of the robot, but instead were asked to indicate to what extent each of the 37 items on the Predisposition to Anthropomorphize Scale was perceived as 'typically human' (i.e. human nature) or 'uniquely human' respectively. Human nature was measured with a 7-point response format ranging from 'not at all' (coded with a 1) to 'very much' (coded with a 7). Human uniqueness was measured with a dichotomous response format with 'not unique' (coded with a 0) and 'unique' (coded with a 1) as options. Both these evaluations took approximately 5 minutes to complete.

4.2.2 Results and discussion

Model test

In this section, the hypotheses were tested that items and persons can be mapped onto a single one-dimensional scale, and that items are invariantly ordered according to the difficulty with which they are ascribed to a robot. Four tests were conducted. First, fit statistics (for an overview, see Bond & Fox, 2013) were used to test whether items and persons fitted the Rasch model. For assessing item fit, infit and outfit were used. Infit indicates unexpected observations on items that are close to a person's predisposition. Outfit indicates unexpected observations on items that are far away from a person's predisposition. Infit MS-values up to 1.20 are considered good, and outfit MS-values up to 1.50 are considered good (Wright, Linacre, Gustafson, & Martin-Lof, 1994). The second test determined whether the items had a sufficient spread across the human-likeness continuum. The third test determined whether the items all belonged to one dimension. The fourth test determined whether items were invariantly ordered according to the difficulty of attributing them to robots.

Item fit Ideally, each of the items should contribute in a meaningful way to the measuring instrument, indicated by a sufficient item fit. Results showed that most items fitted the model sufficiently with infit MS values ≤ 1.20 and outfit MS values ≤ 1.50 (see Table 4.1 for estimated item difficulties), except for items 1 ('experience pain', MS outfit = 2.09), 18 ('jump', MS infit = 1.22), 26 ('anticipate on surroundings', MS outfit = 1.62), 32 ('organized', MS outfit = 2.74), 33 ('estimate distances', MS outfit = 2.12), and 37 ('avoid objects', MS outfit = 1.97).

Table 4.1: Item difficulties (δ), infit- and outfit mean squares, human uniqueness and human nature of the Predisposition to Anthropomorphize Scale in Study 5.

	Item	δ (SE)	Infit MS	Outfit MS	Human uniqueness	Human nature
1.	Experience pain	4.08 (.60)	1.04	2.04	0.00	0.54
2.	Unhappy about the dilemma	3.77 (.52)	0.91	0.69	0.88	0.75
3.	Imaginative	3.52 (.47)	0.86	0.51	0.71	0.78
4.	Angry	3.32 (.44)	0.87	0.71	0.06	0.78
5.	Empathize	2.84 (.36)	0.73	0.46	0.41	0.73
6.	Happy	2.71 (.35)	0.88	0.86	0.12	0.69
7.	Chose the dilemma	2.71 (.35)	1.03	1.13	0.82	0.75
8.	Satisfied	2.29 (.30)	0.79	0.58	0.12	0.54
9.	Responsible	2.29 (.30)	0.86	0.64	0.24	0.72
10.	Free will	2.20 (.30)	1.14	1.14	0.18	0.63
11.	Understand others' emotions	1.88 (.27)	0.86	0.62	0.24	0.66
12.	Ambitious	1.68 (.26)	0.95	0.72	0.76	0.81
13.	Understands the dilemma	1.21 (.23)	1.04	0.92	0.94	0.67
14.	Recognize others' emotions	0.55 (.21)	0.89	0.85	0.06	0.60
15.	Intention not to harm others	0.47 (.21)	1.06	1.03	0.59	0.65
16.	Think about the dilemma	0.42 (.21)	0.94	0.85	0.88	0.81
17.	Self-conscious	0.42 (.21)	0.99	0.95	0.29	0.73
18.	Jump	0.26 (.20)	1.22	1.24	0.00	0.34
19.	Deliberate actions	-0.15 (.20)	1.11	1.38	0.06	0.58
20.	Talk	-0.19 (.20)	0.95	1.04	0.24	0.78
21.	Solve riddles	-0.47 (.20)	1.01	0.99	0.47	0.67
22.	Recognize voices	-0.72 (.21)	1.11	1.46	0.00	0.48
23.	Understand language	-0.89 (.21)	0.86	0.75	0.24	0.75
24.	Rational	-0.93 (.21)	1.05	1.36	0.71	0.76
25.	See depth	-0.97 (.21)	1.07	1.21	0.00	0.55
26.	Anticipate on surroundings	-1.45 (.23)	1.09	1.62	0.12	0.58
27.	Conscious about surroundings	-1.66 (.24)	0.72	0.54	0.00	0.50
28.	Detect color	-1.83 (.24)	0.92	0.87	0.00	0.54
29.	Purposeful	-1.96 (.25)	0.92	1.50	0.18	0.58
30.	Calculate	-2.16 (.26)	0.94	0.78	0.35	0.67
31.	See	-2.54 (.29)	1.13	0.98	0.00	0.35
32.	Organized	-2.73 (.31)	1.17	2.74	0.18	0.52
33.	Estimate distances	-2.73 (.31)	1.12	2.12	0.06	0.48
34.	Pick up objects	-3.05 (.34)	1.02	0.86	0.06	0.54
35.	Walk	-3.61 (.42)	1.02	1.49	0.00	0.48
36.	Detect objects	-3.02 (.48)	1.07	0.99	0.00	0.49
37.	Avoid objects	-4.60 (.61)	0.99	1.97	0.00	0.39

Considering the notion that the Rasch model is stochastic and that data depend on probability (and not on certainty), some misfit is to be expected (Shaw, 1991). An acceptable six of the 37 items had MS infit or outfit values outside of the proposed boundaries.

Item difficulties were estimated with a reliability of α = .98, and the average item difficulty was anchored at M = .00 logits (SD = 2.34, Range = -4.60 to 4.08). Infit MS values of the 37 items ranged from 0.72 to 1.22 (M = 0.98, SD = 0.12), and outfit MS values ranged from 0.46 to 2.74 (M = 1.10, SD = 0.50).

Person fit The purpose of person fit measurement is to detect response patterns that are unlikely given the model (Meijer & Sijtsma, 2001). More specifically, person fit indicates whether a person's responses are likely given his/her individual predisposition to anthropomorphize. Individual predispositions to anthropomorphize were estimated with a reliability of α = .80. The average predisposition was M = -.21 logits (SD = 1.16; Range = -4.13 to 2.74). For a reasonable ten out of 124 participants (8.1%), the model prediction did not fit the data as indicated by a t-value of $t \geq 1.96$.

Item spread All items and persons are mapped onto a single scale in Figure 4.1. As can be seen in this figure, the spread of items amongst the human-likeness continuum sufficiently covers the spread of persons. In other words, the current scale was able to reliably measure individual predispositions to anthropomorphize for all participants in the current sample. It also appeared that the top region of the scale comprised many items, but not so many persons. Thus, some items appeared to be too difficult for persons in the present sample to attribute to the robot, and therefore did not contribute to the assessment of individual differences in people's predisposition to anthropomorphize. For this reason, some of these items were deleted from the scale in follow-up studies.

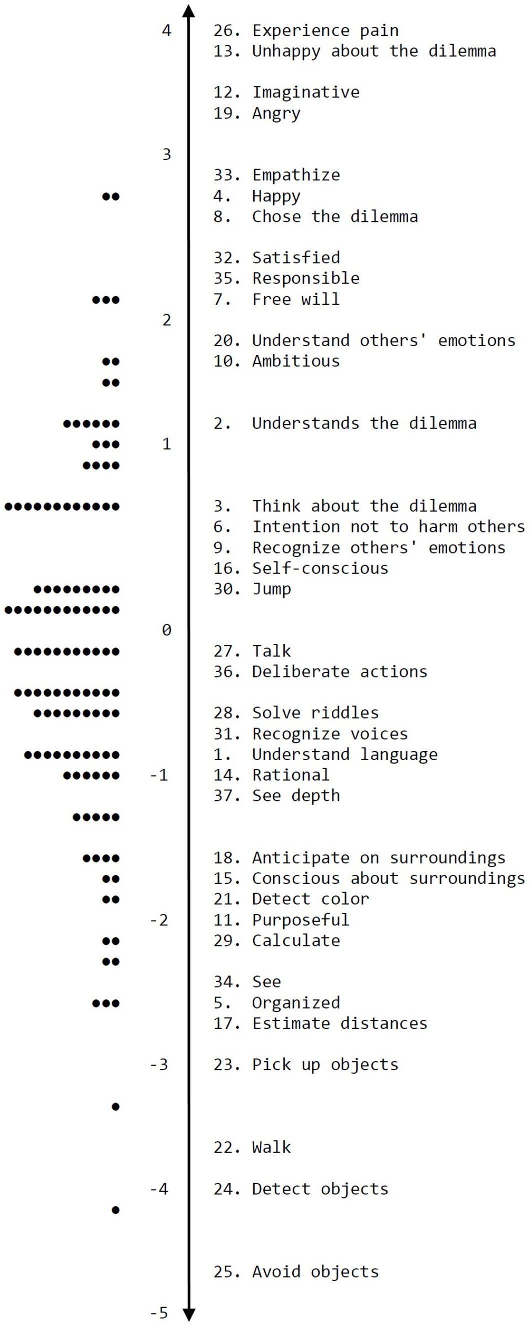


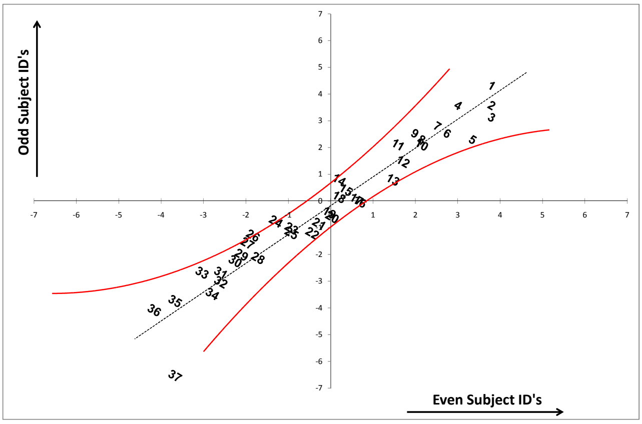
Figure 4.1: Item-person map of Study 5, displaying the estimates of participant's predisposition to anthropomorphize and the item difficulty linked with each human-like characteristic mapped onto a single scale of equal additive units. Each number represents an item. Each dot represents a person.

Dimensionality Next, the expected unidimensionality of the Predisposition to Anthropomorphize Scale was tested. Results showed that the Rasch model explained 52.8% of the variance in the data (for computational details, see Linacre, 2003). A Principal Component Analysis was performed on the standardized residuals (i.e. the data not explained by the model), which checks whether multiple items share the same unexpected response pattern (for details, see Smith Jr, 2002). If the model would fit perfectly, then 52.5% of the overall variance would be quantification variance, revealing a slight overfit to the model. Because the Rasch model estimates probabilities for discrete events (i.e., whether a person attributes a certain human-like characteristic to a robot or not), substantial quantification variance is to be expected (see also Haans et al., 2012). The empirical proportion of unexplained variance (i.e., 47.2%) was thus highly similar to the proportion of quantification variance one would expect with a perfect data-to-model fit.

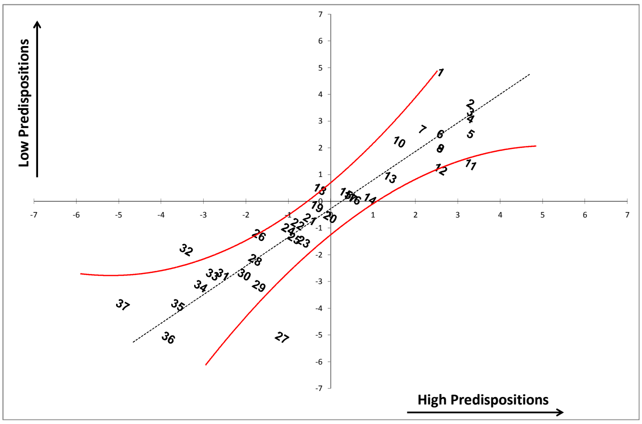
An additional factor would only result in an increase of a trivial 6.8% in the proportion of explained variance, and the set of items thus largely tapped into a single factor only. These findings supported the expected uni-dimensionality of the scale, and showed that individual differences in predispositions to anthropomorphize could be assessed on a single scale of equal additive units. In other words, all human-like characteristics included in the Predisposition to Anthropomorphize Scale were successfully mapped onto a single dimension, ranging from low to high human-likeness.

Invariant ordering To test the hypothesis that items on the Predisposition to Anthropomorphize Scale were invariantly ordered, the sample was split in half and item difficulties were estimated twice: once for participants with even and once for participants with odd identification numbers. Consistent with the hypothesis of person-independent item difficulties, the two estimates of the 37 items were highly similar, $r =$

.97, $p < .001$. The item invariance plot is provided in Figure 4.2a. As can be seen in this figure, the ordering of items on the Predisposition to Anthropomorphize Scale by their difficulty to ascribe them to the robot is similar across the samples of participants with even and odd identification numbers.



(a) even vs odd subject id's



(b) high vs low predispositions

Figure 4.2: Item invariance plots of the item difficulties of subjects with (a) even and odd identification numbers and (b) high and low predispositions to anthropomorphize in Study 5. Each number represents an item, corresponding with the numbers in Table 4.1.

We also performed the 'Wright's challenge' (see Bond & Fox, 2013). For this, the sample was split in half once more, but this time according to the participants' estimated predispositions. More specifically, item difficulties were estimated for participants with high predispositions and for participants with low predispositions to anthropomorphize separately. The estimates of the 37 items were again highly similar, $r = .92$, $p < .001$. The item invariance plot is provided in Figure 4.2b. As can be seen in this figure, the ordering of items on the Predisposition to Anthropomorphize Scale is also similar across the samples of participants with high and low individual predispositions to anthropomorphize.

Construct validity

To test the hypothesis that the ordering of items of the Predisposition to Anthropomorphize Scale according to their difficulty was related to perceived human nature and human uniqueness, item difficulties were compared with the mean scores on human nature and human uniqueness. Results showed significant correlations between item difficulties and human nature ($r = .60$, $p < .001$) and between item difficulties and human uniqueness ($r = .44$, $p < .01$). The higher a characteristic was rated on human nature and/or human uniqueness by participants in the content validity groups, the more difficult it was for participants in the survey to attribute that specific human-like characteristic to a robot. These results support the expectation that the difficulty to attribute a specific characteristic to a robot (δ_i) is related to that characteristic's perceived human nature and human uniqueness.

Additionally, a significant correlation was found between human nature and human uniqueness ($r = .71$, $p < .001$), which indicates that items that were rated high on human nature were also more likely to be indicated as being uniquely human, and that items that were rated low on human nature were less likely to be indicated as being uniquely human. This finding supports the expectation that all human-like characteristics can be mapped onto a one-dimensional scale.

4.2.3 Conclusions

In the current study, a 37-item Predisposition to Anthropomorphize Scale was developed and tested on its construct validity. Results showed that people's responses sufficiently fitted the Rasch model, indicated by an acceptable data-to-model fit. These results supported the expectation that human-like characteristics can be invariantly ordered with respect to the probability of ascribing them to robots. The Predisposition to Anthropomorphize Scale also appeared to cover a wide range of the human-likeness continuum. However, some of the items on the scale had such a high difficulty that they did not contribute to the assessment of individual differences in predisposition to anthropomorphize. Those items will be deleted in the next study.

As expected, uniquely human characteristics were shown to be more difficult to attribute to robots than non-unique ones, as was indicated by the significant correlations between human uniqueness and item difficulties. Additionally, items high in human nature were found to be more difficult to attribute to a robot than those low in human nature. These results indicated that the difficulty of attributing a specific item of the Predisposition to Anthropomorphize Scale to a robot was related to that item's perceived human-likeness. The scale thus appeared to have high construct validity. In the next study, the Predisposition to Anthropomorphize Scale's convergent validity will be tested by comparing it with existing measuring instruments.

4.3 Study 6

In this study, the Predisposition to Anthropomorphize Scale was tested on its convergent validity. Convergent validity refers to the extent to which estimates obtained with the scale converged with those obtained with two commonly used measuring instruments for anthropomorphism:

the Waytz-instrument and the Godspeed-instrument. We hypothesized that estimates made with the three different instruments would be related, indicating high convergent validity. In addition, two different robots were evaluated on their perceived human-likeness. An invariant ordering of the items on the Predisposition to Anthropomorphize Scale for those two robots was expected.

Since the Godspeed- and Waytz-instruments only focus on subsets of human-like characteristics (i.e., appearances and thoughts respectively), they might be less effective in differentiating between a wide range of predisposition levels than the Predisposition to Anthropomorphize Scale. Such differences in range could attenuate the correlations between the measures (see e.g., Hunter & Schmidt, 2004). To investigate the extent to which the three measuring instruments relate to each other, items of the three instruments were mapped together onto a single dimension.

4.3.1 Method

Participants and design

One hundred and thirty one participants sampled through social media participated in the current study. Of these 131 participants, 48 were male and 83 female ($M_{age} = 34.86$, $SD_{age} = 17.59$, Range = 13 to 77; two participants did not indicate their age). Participants were randomly assigned to one of two experimental conditions in which they watched a video of a robot that either resembled human-like physical features ($n = 68$) or that resembled human-like cognitive features ($n = 63$). The two robots did not differ on any of the three included anthropomorphism measuring instruments (all t's < 1.31, all p's $> .20$), allowing data of both experimental conditions to be combined into a single sample for the analyses. All participants participated voluntarily, gave informed consent, and were not compensated for participation.

Materials and procedure

Participants performed the study online. On the welcome page, participants could choose to complete the study in Dutch or in English, after which they were provided information about the procedure of the study in their preferred language. Next, they watched a short (about 1 minute) video of one of the two robots, depending on the experimental condition they were in. The robot with human-like physical features was running around and capable of pouring water in a cup, and the robot with human-like cognitive features appeared to become angry at a person who left dirt on the floor.

After participants watched the video of one of the two robots, they completed three measuring instruments for anthropomorphism. The first one was a 25-item version of the Predisposition to Anthropomorphize Scale that was developed and tested in Study 5 and adjusted for the current study. Some of the most difficult items, as well as the easiest one, were deleted because they were expected to contribute little to estimations of people's predisposition to anthropomorphize (i.e. items 1, 3, 6, 7, 8, 10, and 37 in Table 4.1). Three items were deleted because the construct validity tests in Study 5 showed that they did not sufficiently relate to the concept (i.e. items 24, 28, and 32, in Table 4.1). Item 15 in Table 4.1 was deleted because it was phrased as a double negation.

The second questionnaire was the Godspeed-instrument, which consisted of 5 items with a 5-point response format[5]. Participants' average responses across the five items were used in the analyses ($\alpha = .71$).

The third questionnaire was the Waytz-instrument, which consisted of 6 items with a 5-point response format. Participants' average responses across the six items were used in the analyses ($\alpha = .78$).

[5]Five dummy items were included to make the goal of this questionnaire less obvious.

After completing the anthropomorphism questionnaires, participants indicated their age and gender, were debriefed and thanked for their participation. The study took approximately 10 minutes to complete.

4.3.2 Results and discussion

Model test

In this section, the hypotheses that items and persons can be mapped onto a single one-dimensional scale, and that items are invariantly ordered according to the difficulty with which they are ascribed to a robot are tested. The section has the same structure as in Study 5.

Item fit As in Study 5, most items fitted the model sufficiently with infit MS values ≤ 1.20 and outfit MS values ≤ 1.50 (see Table 4.2 for estimated item difficulties), except for items 13 ('understands the dilemma', MS outfit = 5.21), 5 ('empathize', MS outfit = 1.64), 4 ('angry', MS outfit = 2.01), 25 ('see depth', MS outfit = 5.97), and 34 ('pick up objects', MS infit = 1.59).

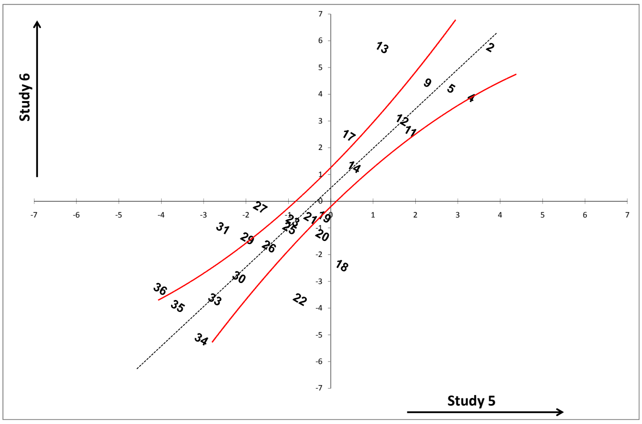
Item difficulties were estimated with a reliability of $\alpha = .98$. The average item difficulty was anchored at $M = .00$ logits ($SD = 3.14$, Range = -5.16 to 5.76). Infit MS values of the 25 items ranged from 0.71 to 1.59 ($M = 1.00$, $SD = 0.18$). Outfit MS values of the 25 items ranged from 0.10 to 5.97 ($M = 1.23$, $SD = 1.35$).

Person fit Individual predispositions to anthropomorphize were estimated with a reliability of $\alpha = .78$. The average predisposition was $M = .33$ logits ($SD = 1.59$, Range = -5.70 to 5.47). For a reasonable eight out of 131 participants (6.1%), the model prediction did not fit the data as indicated by a t-value of $t \geq 1.96$.

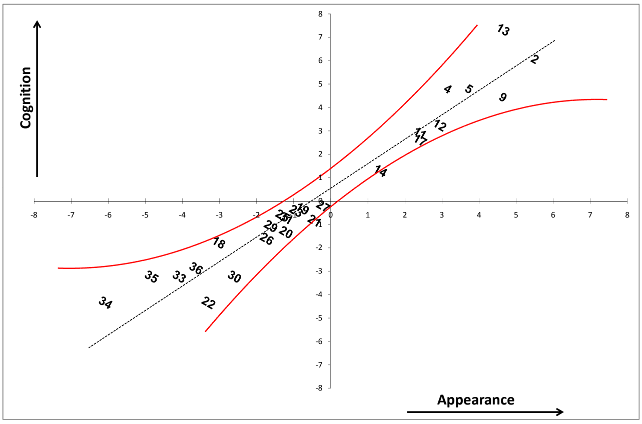
Table 4.2: Item difficulties (δ), infit- and outfit mean squares of the Predisposition to Anthropomorphize Scale in Study 6.

	Item	δ (SE)	Infit MS	Outfit MS
13.	Understands the dilemma	5.76 (.76)	1.30	5.21
2.	Unhappy about the dilemma	5.76 (.76)	0.71	0.10
9.	Responsible	4.43 (.46)	0.76	0.22
5.	Empathize	4.22 (.43)	1.15	1.64
4.	Angry	3.88 (.39)	1.00	2.01
12.	Ambitious	3.05 (.31)	1.03	0.73
11.	Understand others' emotions	2.62 (.28)	0.79	0.81
17.	Self-conscious	2.46 (.27)	1.04	0.97
14.	Recognize others' emotions	1.29 (.22)	1.03	0.97
27.	Conscious about surroundings	-0.23 (.21)	0.93	0.77
19.	Deliberate actions	-0.58 (.21)	1.08	1.07
21.	Solve riddles	-0.63 (.21)	1.20	1.17
23.	Understand language	-0.72 (.22)	1.04	0.89
31.	See	-1.01 (.22)	0.94	0.92
25.	See depth	-1.01 (.22)	1.01	5.97
20.	Talk	-1.27 (.24)	0.96	0.95
29.	Purposeful	-1.39 (.24)	0.94	0.68
26.	Anticipate on surroundings	-1.69 (.26)	0.92	0.72
18.	Jump	-2.40 (.31)	1.01	0.92
30.	Calculate	-2.84 (.36)	0.92	0.81
36.	Detect objects	-3.28 (.41)	0.82	0.31
33.	Estimate distances	-3.67 (.48)	0.78	0.51
22.	Recognize voices	-3.67 (.48)	1.06	0.90
35.	Walk	-3.92 (.52)	1.12	0.83
34.	Pick up objects	-5.16 (.82)	1.59	0.70

Dimensionality The Rasch model explained 63.7% of the variance in the data. If the model would fit perfectly, then 63.5% of the overall variance would be quantification variance. The empirical proportion of unexplained variance (i.e., 36.3%) was thus highly similar to the proportion of quantification variance expected with a perfect data-to-model fit. An additional factor would only result in an increase of a trivial 3.6% in the proportion of explained variance, and the set of items thus largely tapped into a single factor only.



(a) Study 5 vs Study 6



(b) appearance vs cognition

Figure 4.3: Item invariance plots of the item difficulties of (a) Studies 5 and 6, and (b) persons in the appearance and cognition conditions in Study 6. Each number represents an item, corresponding with the numbers in Table 4.2. Red lines indicate 95% confidence intervals.

Invariant ordering The ordering of the item difficulties was highly similar to that obtained in Study 5, as indicated by a strong positive correlation between the item difficulties estimated in Studies 5 and 6 (r = .88, p < .001, see Figure 4.3a for the invariance plot). This result supports the expectation that the probability with which the various human-like characteristics are ascribed to robots is largely independent of the individual's predisposition to do so. In other words, the scale showed an ordering of human-like characteristics that is similar for different individuals in different samples.

To explore whether the expected invariance of item difficulties also holds across the two different robots that were evaluated, the sample was split in half with respect to the robot a participant evaluated. Consistent with the hypothesis of person-independent item difficulties, the two sets of estimates (one for the robot with physical features, the other for the robot with cognitive features) were highly similar (r = .97, p < .001, see Figure 4.3b for the invariance plot). This finding again supports the expectation that all human-like characteristics are invariantly ordered.

Convergent validity

To test whether estimates obtained with the Predisposition to Anthropomorphize Scale converged with the two commonly used measuring instruments for anthropomorphism, the three scales were compared. Results indicated a low, but statistically significant correlation between the Predisposition to Anthropomorphize Scale and the Godspeed-instrument (r = .22, p = .01). After correcting for measurement error attenuation, the correlation remained rather low (r = .29, for computational details, see e.g., Charles, 2005). In addition, a moderate and statistically significant correlation was found between the Predisposition to Anthropomorphize Scale and the Waytz-instrument (r = .46, p < .001). After correcting for measurement error attenuation, this correlation remained rather moderate (r = .59). Thus, the Predisposition to Anthropomorphize Scale converged to some extent with both the Godspeed- and Waytz-instruments.

In order to compare the three measuring instruments on their effectiveness in differentiating between different levels of anthropomorphism, the items of the Waytz- and Godspeed-instruments were mapped onto the Predisposition to Anthropomorphize Scale (see Figure 4.4). Since the items of the Godspeed- and Waytz-instruments had 5-point response formats, the partial credit Rasch model was used (for details, see Bond & Fox, 2013). This model splits non-dichotomous questions into multiple 'steps', which are the transitions from one response category to the next. For example, 'W1.1' indicates the transition between response categories '1' and '2' of the first item of the Waytz-instrument.

As can be seen in the figure, several items of the Godspeed- and the Waytz-instruments were disordered (for example, $\delta_{\mathrm{W5.1}} > \delta_{\mathrm{W5.2}}$ and $\delta_{\mathrm{G4.1}} > \delta_{\mathrm{G4.2}}$). Disordering of items implies that less frequently observed response categories do not contribute to the measurement (for details about disordering of items, see Linacre, 2001). Thus, some of the items in the Waytz- and Godspeed-instruments had response categories that were not used and thus did not contribute to the measurement of anthropomorphism.

In addition, items of the Waytz-instrument appeared to cover only a limited part of the human-likeness continuum (indicated by the blue region that covers all items starting with a 'W'). In particular, the Waytz-instrument largely contained human-like characteristics that were difficult to attribute to robots. The items of the Godspeed-instrument appeared to cover a wider and more intermediate range of the human-likeness continuum (indicated by the red region that covers all items starting with a 'G'). The items of the Predisposition to Anthropomorphize Scale, however, covered the widest range of the human-likeness continuum and could thus be expected to differentiate better amongst the range of people's predispositions to anthropomorphize.

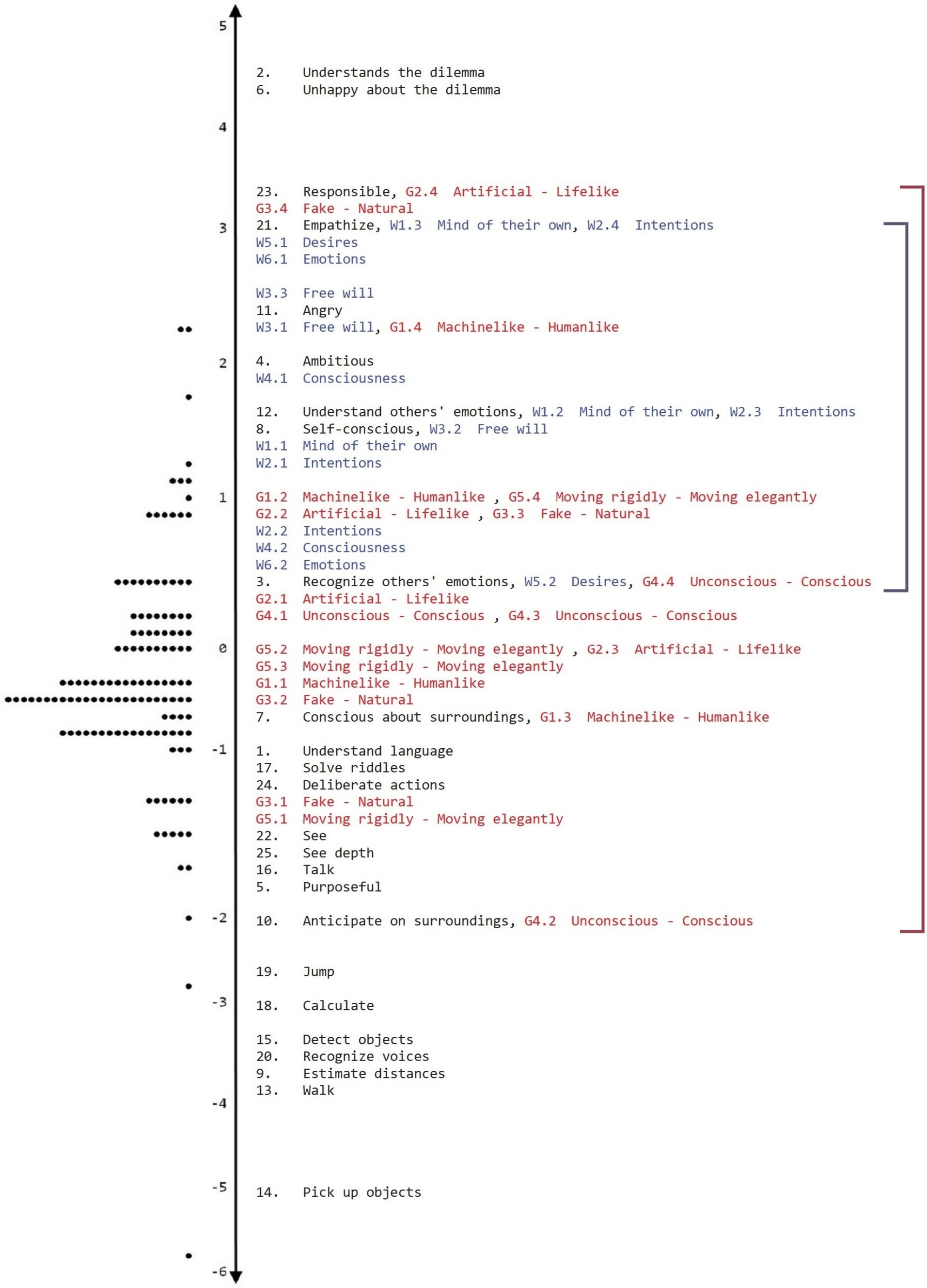


Figure 4.4: Items of the Predisposition to Anthropomorphize Scale, the Godspeed-instrument (items starting with a G), and the Waytz-instrument (items starting with a W) in Study 6 mapped onto a single scale. Each number represents an item or a Rasch-Andrich threshold (see Andrich, 1982 for details). Each dot represents a person.

4.3.3 Conclusions

In the current study, an adjusted 25-item version of the Predisposition to Anthropomorphize Scale was compared with two available measuring instruments for anthropomorphism to test for convergent validity. Results showed that, as in Study 5, people's responses sufficiently fitted the Rasch model, indicated by an acceptable data-to-model fit. This result supported the expected invariant ordering of human-like characteristics with respect to their item difficulty. Additionally, convergent validity was found. That is, the Predisposition to Anthropomorphize Scale correlated well with available measuring instruments for anthropomorphism, indicating that the three measuring instruments for anthropomorphism were related to each other.

The Predisposition to Anthropomorphize Scale also covered a wider range of the human-likeness continuum than the Waytz- and God-speed-instruments did. This finding could explain the low to moderate correlations between the different instruments, because the two existing instruments appeared to measure smaller ranges of the human-likeness continuum than the Predisposition to Anthropomorphize Scale. People with very high and very low predispositions to anthropomorphize cannot sufficiently be distinguished from those with moderate predispositions to anthropomorphize by the Waytz- and Godspeed-instruments, and could thus have a negative influence on the correlations between the measurement instruments.

No differences in predispositions to anthropomorphize were found between the two robots used in the study, so they were perceived as equally human-like. This finding was comparable to that in Study 4 (see Chapter 3), in which no difference between artificial agents that showed consistent versus inconsistent social cues was found on a tentative version of the scale. An interesting question thus is how well the Predisposition to Anthropomorphize Scale differentiates between types of agents that are expected to be different in their perceived human-likeness. In the next study, people's predispositions to anthropomorphize humans, robots, computers, and algorithms will be compared.

4.4 Study 7

In this study, the Predisposition to Anthropomorphize Scale's was tested on its sensitivity for differentiating between agent types. Sensitivity was tested by comparing people's responses to four different kinds of agents: humans, robots, computers, and algorithms. We hypothesized that the Predisposition to Anthropomorphize Scale would successfully differentiate humans from other types of agents; with participants being more inclined to attribute human-like characteristics to robots than to computers, and in turn, attribute more human-like characteristics to computers than to algorithms.

The Waytz-instrument was included in the design of the current study, enabling us to compare the Predisposition to Anthropomorphize Scale's sensitivity with that of the Waytz-instrument and demonstrate the benefits of the Rasch-based instrument. Since the Waytz-instrument included mainly items about the more difficult to ascribe human-like characteristics (see Study 6), it was expected to be relatively insensitive to differentiate between the perceived human-likeness of humanoid robots, computers, and algorithms, compared to the Predisposition to Anthropomorphize Scale.

4.4.1 Method

Participants and design

Two hundred and two participants (89 males and 113 females; M_{age} = 34.69, SD_{age} = 11.12, Range = 18 to 76) participated in an online experiment and were randomly assigned to one of four experimental conditions (Agent type: human vs. robot vs. computer vs. algorithm) of a between-subjects design. Participants (85% American) were recruited via Amazon Mechanical Turk (MTurk), and were paid $1 for their participation. The experiment took approximately 6 minutes to complete.

Materials and procedure

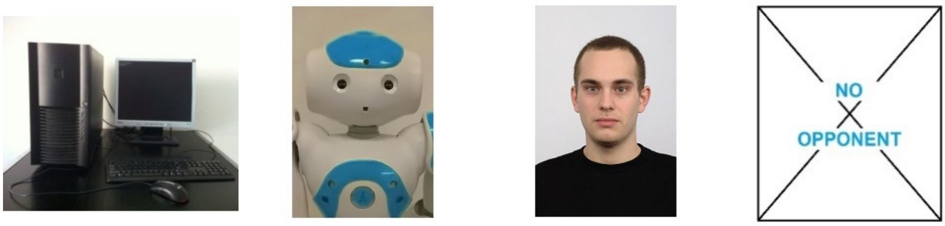
Participants performed the study online. To create a social interaction that provides opportunities to anthropomorphize the agents, participants played the Ultimatum Game (see e.g., Güth et al., 1982). In this game, two players divide a sum of credits. The first player (the agent) proposes a certain division and the second player (the participant) decides whether he/she accepts or rejects the offer. Participants in the algorithm condition were told that there were no other players available, and that they would be connected to an algorithm that would 'randomly generate offers' during the game. Participants in the other three conditions were told to be playing the game with humans, robots, or computers. During the game, participants were shown pictures of the other players. For an example of each of the agent types, see Figure 4.5.

After playing the Ultimatum Game, participants completed a 5-item[6] version of the Waytz-instrument used in Study 6 ($\alpha = .97$) and an adjusted 19-item version of the Predisposition to Anthropomorphize Scale. These 19 items were selected from the original set of 37 items to create a short questionnaire while still covering a wide range of the human-likeness continuum.

After completing the anthropomorphism questionnaire, participants indicated their age and gender, were debriefed, thanked for their partici-

[6] Due to the nature of the Ultimatum Game, the 'desires' item from the Waytz-instrument was not used in the current study.

pation, and paid through the MTurk system. The study took approximately 6 minutes to complete.



(a) (b) (c) (d)

Figure 4.5: Examples of pictures used in the (a) computer, (b) robot, (c) human, and (d) algorithm conditions.

4.4.2 Results and discussion

Model test

In this section, the hypotheses that items and persons can be mapped onto a single one-dimensional scale, and that items are invariantly ordered according to the difficulty with which they are ascribed to a robot are tested. The section has the same structure as in Studies 5 and 6.

Item fit As in Studies 5 and 6, most items fitted the model sufficiently with infit MS values ≤ 1.20 and outfit MS values ≤ 1.50 (see Table 4.3 for estimated item difficulties), except for items 14 ('recognize others' emotions', MS infit = 1.29), 12 ('ambitious', MS outfit = 1.59), 33 ('estimate distances', MS outfit = 1.97), 21 ('solve riddles', MS infit = 1.34), and 30 ('calculate', MS outfit = 5.59).

Item difficulties were estimated with a reliability of $\alpha = .98$. The average item difficulty was anchored at $M = .00$ logits ($SD = 1.83$, Range = -4.15 to 3.05). Infit MS values of the 19 items ranged from 0.76 to 1.29 ($M =$

1.00, SD = 0.17). Outfit MS values of the 19 items ranged from 0.46 to 5.59 (M = 1.14, SD = 1.12).

Table 4.3: Item difficulties (δ), infit- and outfit mean squares of the Predisposition to Anthropomorphize Scale in Study 7.

	Item	δ (SE)	Infit MS	Outfit MS
14.	Recognize others' emotions	3.05 (.36)	1.29	1.03
2.	Unhappy about the dilemma	2.36 (.32)	1.10	0.70
4.	Angry	2.17 (.31)	0.83	0.60
11.	Understand others' emotions	1.90 (.29)	0.93	0.53
13.	Understands the dilemma	1.51 (.27)	0.99	0.79
27.	Conscious about surroundings	1.04 (.25)	0.78	0.46
18.	Jump	1.04 (.25)	0.86	0.73
35.	Walk	0.27 (.22)	0.87	0.64
17.	Self-conscious	0.17 (.22)	0.76	0.54
26.	Anticipate on surroundings	0.12 (.22)	0.96	0.64
34.	Pick up objects	-0.02 (.22)	0.92	0.71
12.	Ambitious	-0.30 (.21)	1.09	1.59
20.	Talk	-0.30 (.21)	0.99	0.77
33.	Estimate distances	-0.95 (.21)	1.05	1.97
36.	Detect objects	-1.20 (.21)	0.80	0.71
21.	Solve riddles	-1.54 (.21)	1.34	1.17
29.	Purposeful	-2.12 (.22)	1.17	1.14
23.	Understand language	-3.04 (.24)	1.03	1.31
30.	Calculate	-4.15 (.29)	1.25	5.59

Person fit Individual predispositions to anthropomorphize were estimated with a reliability of α = .88. The average predisposition was M = -.15 logits (SD = 3.28, Range = -5.79 to 5.31). For a reasonable ten out of 202 participants (5.0%), the model prediction did not fit the data as indicated by a t-value of $t \geq 1.96$.

Dimensionality The Rasch model explained 52.4% of the variance in the data. If the model would fit perfectly, then 52.1% of the overall variance would be quantification variance. As in Studies 5 and 6, the proportion of unexplained variance (i.e., 47.6%) was highly similar to the proportion of quantification variance expected with a perfect data-to-model fit. An additional factor would result in an increase of 8.3% of the explained variance, and the set of items thus largely tapped into a single factor only.

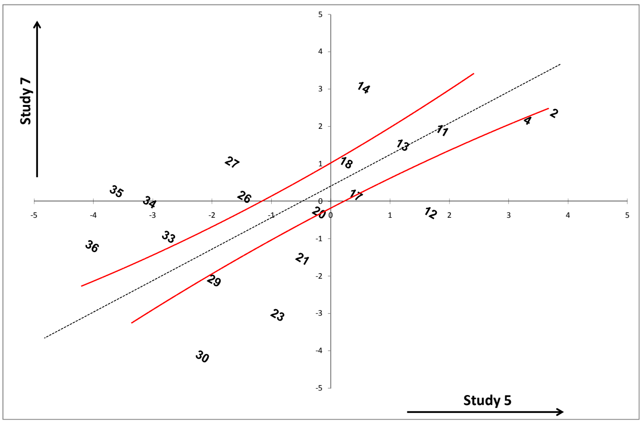


Figure 4.6: Item invariance plot of the item difficulties in Studies 5 and 7. Each number represents an item. Red lines indicate 95% confidence intervals.

Next, the ordering of items in the current study was compared with that of Study 5. Results showed a moderate but significant correlation between the two estimates ($r = .58$, $p < .01$, see Figure 4.6 for the invariance plot). Despite the significant correlation between the item difficulties, many of the estimated difficulties appeared outside of the 95% confidence interval, indicating substantial differences between the two studies. Some of these items (i.e. 'pick up objects', or 'walk') had lower item difficulties in Study 5 than in Study 7, and some (i.e. 'calculate', or 'understands language') had higher item difficulties in Study 5 than in Study 7. Some human-like characteristics were thus more easy or more difficult to attribute to a specific type of agent than other characteristics. More specifically, the expected invariant ordering of human-like characteristics with respect to their probability of being ascribed to non-human agents was not supported when other agents than robots were evaluated. This finding will be investigated in more detail in the following sections.

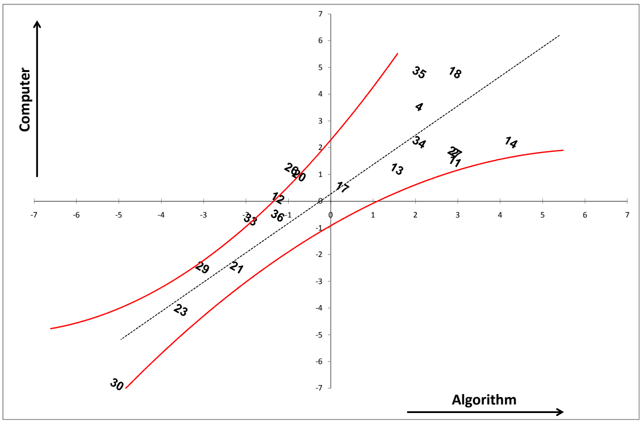
Variations in item difficulties

To explore which human-like characteristics may be different in their probability of being attributed to humans, robots, computers, and algorithms, item difficulties were estimated separately for each of these four agent types. Although all correlations between the four sets of item difficulties were significant (see Table 4.4), some human-like characteristics were clearly different in their probability of being ascribed to these agents.

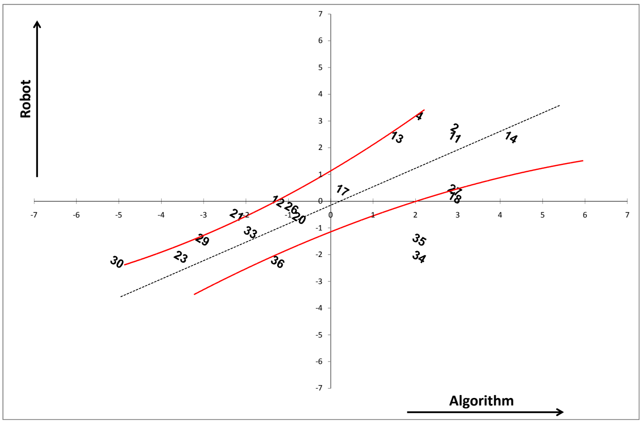
Table 4.4: Correlations between item difficulties in the experimental conditions in Study 7.

	Human		Robot		Computer	
Algorithm	$r = .74$	$p < .001$	$r = .68$	$p < .01$	$r = .88$	$p < .001$
Computer	$r = .59$	$p < .01$	$r = .50$	$p = .03$		
Robot	$r = .73$	$p < .001$				

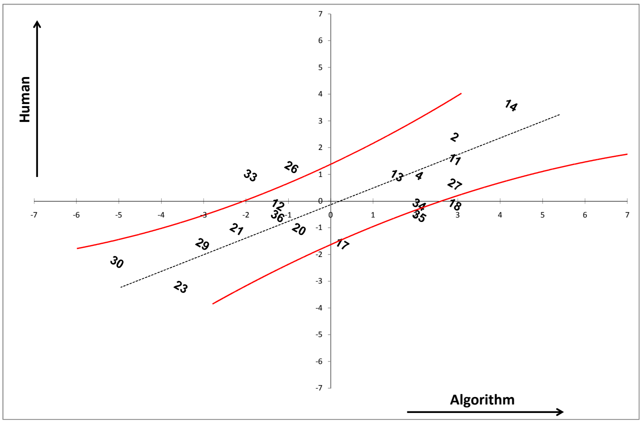
As can be seen in the plots in Figure 4.7, three of the 19 items (e.g., item 18 'jump', 34, 'pick up objects', and 35 'walk') appeared to be less easily ascribed to algorithms and computers than to robots and humans. This should not come as a surprise, as computers and algorithms lack the morphology that accommodates such physical activities.



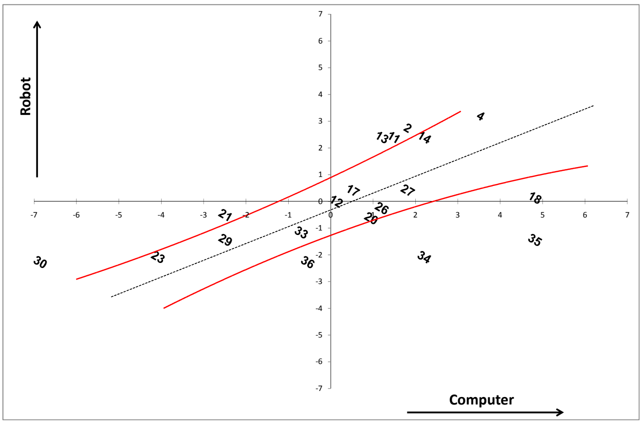
(a) algorithm vs computer



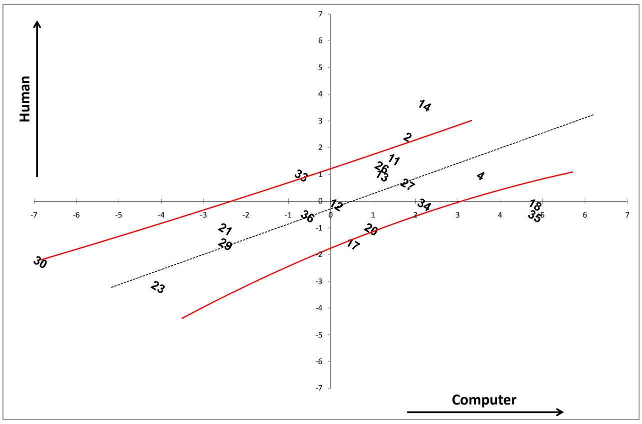
(b) algorithm vs robot



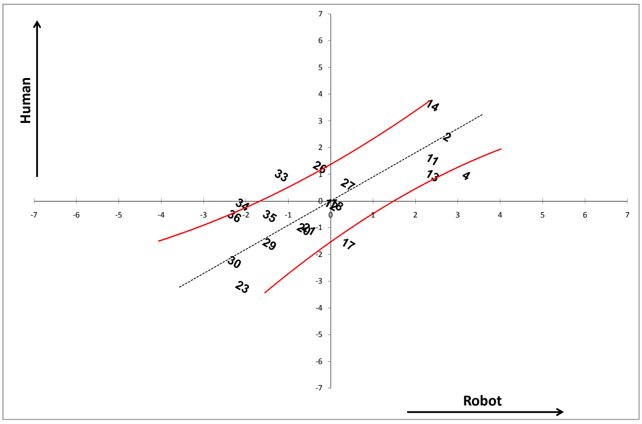
(c) algorithm vs human



(d) computer vs robot



(e) computer vs human



(f) robot vs human

Figure 4.7: Item invariance plots of the item difficulties of each of the four experimental conditions in Study 7: algorithm, computer, robot, and human. Each number represents an item. Red lines indicate 95% confidence intervals.

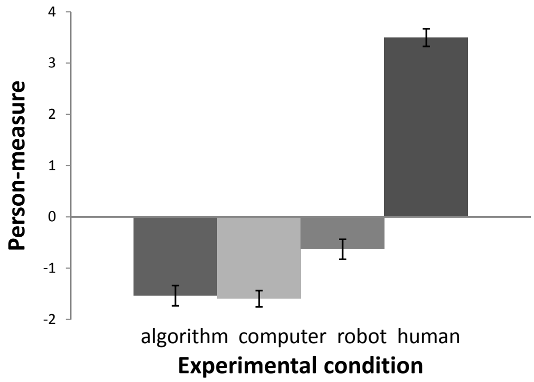
This finding revealed the importance of having an item bank and carefully selecting the specific human-like characteristics to be included in the scale when comparing different types of agents. Items about physical features might be unfair to computers when comparing their anthropomorphism with that of robots in a similar way as it would be unfair to judge people who are bound to a wheelchair as less human than their able counterparts for not being able to jump.

Sensitivity in differentiating between agents

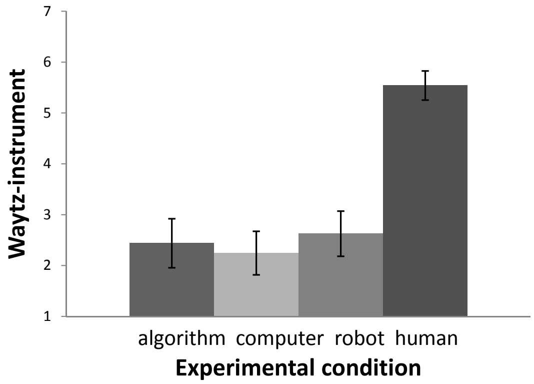
To test the hypothesis that the Predisposition to Anthropomorphize Scale differentiates humans from other types of agents, a one-way Analysis of Variance (ANOVA) was conducted with Agent type (i.e. algorithm, computer, robot, and human) as independent variable and the individual predisposition to anthropomorphize as dependent variable. Results indicated a statistically significant effect of Agent type, $F(3, 201) = 42.04$, $p < .001$, $\eta^2 = .39$, see Figure 4.8a. More specifically, anthropomorphism was highest for humans ($M = 3.50$, $SE = .34$), followed by robots ($M = -0.63$, $SE = .39$), algorithms ($M = -1.53$, $SE = .40$) and computers ($M = -1.60$, $SE = .32$).

Pairwise comparisons (LSD) showed a statistically significant difference between humans and all other agents, $t(198) = 11.03$, $p < .001$. Additionally, marginally significant differences were found between algorithms and robots ($t(198) = 1.78$, $p = .08$) and between computers and robots ($t(198) = 1.90$, $p = .06$).

Interestingly, after removing the three morphology-related items (i.e. items 18, 34, and 35) from the scale, the differences in anthropomorphism between algorithms and robots and between computers and robots became insignificant (both t's < 1.01, both p's $> .32$). The pattern of anthropomorphism remained similar, with estimations being highest for humans ($M = 3.48$, $SE = .34$), followed by robots ($M = -0.83$, $SE = .40$), algorithms ($M = -1.32$, $SE = .40$) and computers ($M = -1.35$, $SE = .33$). The remaining 16 items apparently were not able to differentiate robots from algorithms or computers. This could be caused by the nature of the experiment, because all agents behaved in the exact same way during the Ultimatum Game. When those agents were subsequently compared with respect to characteristics that they were equally likely to possess, it should not come as a surprise that no differences between those agents were found. Nevertheless, this finding is an important one for further development of the Predisposition to Anthropomorphize Scale and will be discussed in more detail later.



(a)



(b)

Figure 4.8: Visualization of participants' (a) person measures on the Predisposition to Anthropomorphize Scale, and (b) scores on the Waytz-instrument per experimental condition in Study 7. Whiskers represent 95% error bars.

To investigate whether the Waytz-instrument was able to differentiate between agent types, a one-way ANOVA was conducted with Agent type (i.e. algorithm, computer, robot, and human) as independent variable and the Waytz-measurement as dependent variable. Results indicated a statistically significant effect of Agent type, $F(3, 201) = 46.49$, $p < .001$, $\eta^2 = .41$, see Figure 4.8b. More specifically, anthropomorphism was highest for humans (M = 5.54, SE = .22), followed by robots (M = 2.63, SE = .22), algorithms (M = 2.44, SE = .24) and computers (M = 2.25, SE = .21).

Pairwise comparisons (LSD) showed a statistically significant difference between humans and all other agents, $t(198) = 11.87$, $p < .001$. This time however, no significant differences were found between either algorithms and robots ($t(198) = .60$, $p = .55$) or computers and robots ($t(198) = 1.22$, $p = .22$).

Finally, a statistically significant correlation was found between measurements obtained with the Predisposition to Anthropomorphize Scale and the Waytz-instrument ($r = .84$, $p < .001$). After correcting for measurement error attenuation, this correlation remained high ($r = .91$).

4.4.3 Conclusions

In the current study, an adjusted 19-item version of the Predisposition to Anthropomorphize Scale was used to investigate the scale's sensitivity, and this sensitivity was compared with that of the Waytz-instrument. Results showed that, as in Studies 5 and 6, people's responses sufficiently fitted the Rasch model, indicated by an acceptable data-to-model fit. This result supported the expected invariant ordering of the human-like characteristics with respect to their item difficulty.

An adequate level of sensitivity was found for the Predisposition to Anthropomorphize Scale. The scale was shown to be able to differentiate between all different agent types, except between algorithms and computers. Presumably the conceptual differences between computers and algorithms were too small to be detected by the current version of the scale.

Additionally, the scale's sensitivity dropped when the three morphology-related items were removed from the analysis, which caused the scale to be similarly insensitive to differentiate between robots and other non-biological agents as the Waytz-instrument. This issue will be discussed in more detail later.

4.5 General discussion

The current research was designed to test the hypotheses that anthropomorphism could be successfully mapped onto a one-dimensional scale, and that human-like characteristics would be ordered in a similar way for all individuals in their encounter with different types of agents. To test the scale's psychometric qualities, data were gathered in three studies. These studies had designs with differing contexts and experimental conditions, and people's predispositions to anthropomorphize were compared in different samples with different types of agents.

In the first study, we hypothesized that items and persons could be mapped onto a single one-dimensional scale, and that items would be invariantly ordered according to the difficulty with which they are attributed to a robot. Additionally, the estimated difficulties with which the 37 characteristics are attributed to a robot were expected to be related to their perceived human nature and human uniqueness. In the second study, we hypothesized that estimates made with three different measuring instruments for anthropomorphism would be related. In the third study, we hypothesized that the Predisposition to Anthropomorphize Scale would successfully differentiate humans from other types of agents, with participants being more inclined to attribute human-like characteristics to robots than to computers and to attribute more human-like characteristics to computers than to algorithms.

All hypotheses were confirmed, and results of the current chapter will be discussed in more detail in the next sections. These sections describe the unidimensionality of anthropomorphism, the differences between various human-like characteristics, and the comparison of the three different measuring instruments.

4.5.1 Unidimensionality of anthropomorphism

Across studies, an invariant ordering of human-like characteristics was found, indicating that this ordering was similar for different people in their encounter with different types of agents. This finding was supported by dimensionality tests that indicated that the data could be represented in a one-dimensional structure. In each of the studies an additional factor would result in only a minor increase in the proportion of explained variance. This result supported the hypothesis that anthropomorphism can be viewed as a one-dimensional construct.

Tests of construct validity showed significant correlations between the item difficulties and their perceived human nature and human uniqueness, supporting the expectation that the scale measures anthropomorphism. Together, these findings indicate that human-like characteristics are ordered in such a way that they range from low to high on *a single dimension*, which is in contrast with views about anthropomorphism being a two-dimensional construct (e.g., Złotowski et al., 2014).

4.5.2 Differences between the characteristics

Findings on some of the characteristics were quite unexpected and need some further consideration. For example, the item 'Experience pain' was rated as extremely low in human uniqueness and as medium in human nature, but it appeared to be the most difficult one to ascribe to a robot. Participants may have argued for themselves that a nervous system is necessary for experiencing pain, which is not unique for humans, but is something that robots clearly do not have.

When morphology-related characteristics were disregarded (i.e the items about a robot being able to jump, to walk, and to pick up objects), no significant differences on anthropomorphism between robots, computers, and algorithms were found. I argue that a measuring instrument for anthropomorphism needs to contain at least some human-like characteristics that a robot does not objectively have, because we are interested in people's *perceptions* of that robot, which may be different from objective descriptions of it. Of course, including morphology-related items when comparing the anthropomorphism of computers versus robots remains rather trivial (despite the finding that on average one in every ten participants did attribute those items to computers and algorithms). It is therefore important to carefully select the items to be used on the scale when comparing different types of agents.

The argument about including characteristics that robots do *and* do not have is in contrast with those of some researchers who argued that measurements of anthropomorphism should only contain characteristics that are distinctively human (Waytz, Cacioppo, & Epley, 2010) or that they should only include human-like characteristics that robots do not have (Zawieska, Duffy, & Strońska, 2012).

4.5.3 Comparison of measuring instruments

The Predisposition to Anthropomorphize Scale was compared with two available measuring instruments for anthropomorphism to test for convergent validity. These available instruments were the Godspeed- and Waytz-instruments. We hypothesized that measurements obtained with the three instruments would be related. This hypothesis was confirmed by significant correlations between the Predisposition to Anthropomorphize Scale and both other instruments. These correlations were however rather low. Further investigation of the relations between the three instruments indicated that the Predisposition to Anthropomorphize Scale measured on a wider range of the human-likeness continuum than the Godspeed- and Waytz-instruments did. This finding indicated that these two measuring instruments only differentiated within a limited range of the human-likeness continuum. The results also revealed that the Predisposition to Anthropomorphize Scale did not fully cover the middle section of the continuum. In this respect, the scale could be improved by including some items to fully cover the human-likeness continuum.

When the Predisposition to Anthropomorphize Scale and the Waytz-instrument were compared on their sensitivity to differentiate between different types of agents, the effect of covering only a limited range of the human-likeness continuum became prevalent. The Predisposition to Anthropomorphize Scale was better able to differentiate between different types of agents than the Waytz-instrument (especially those agents that were low in human-likeness). This finding confirmed the hypothesis that the Predisposition to Anthropomorphize Scale is sensitive to detect differences in types of agents.

4.5.4 Limitations and future research

The Predisposition to Anthropomorphize Scale was used mainly to test its psychometric properties, and all items were included in almost all analyses within the studies, even when some items did not sufficiently fit the Rasch model. When using the scale as a dependent variable in future experimental studies, misfitting items –if there are any– should be disregarded before doing the final analysis. The Rasch model allows researchers to select items from an item bank without influencing the measurement, providing an item-free assessment of anthropomorphism. Therefore, this method is easily adjustable for different goals, for example when investigating anthropomorphism of different agents than robots (e.g., animals or natural phenomena).

In all studies, the items of the Predisposition to Anthropomorphize

Scale were ordered alphabetically, which may have influenced people's responses to certain items because of order effects. The items on the scale should be ordered randomly in future studies to prevent the occurrence of such ordering effects.

All studies were performed with mainly Dutch and American participants, so their cultural backgrounds and experiences with technology could have been quite similar. Earlier work has shown that people's tendencies to attribute human-likeness to non-humans are related to religion (e.g., Guthrie, 1993), and that people from different countries (such as individualistic versus collectivistic ones) respond differently to computers (Katagiri, Nass, & Takeuchi, 2001) and evaluate robots differently (Shibata, Wada, Ikeda, & Sabanovic, 2009). It would be interesting to investigate whether cultural differences influence people's predispositions to anthropomorphize.

4.5.5 Conclusions

Overall, results of the current chapter are promising and adding to the body of literature in the domain of social responses to robots and artificial agents. The Predisposition to Anthropomorphize Scale provides a reliable way of measuring anthropomorphism, because it measures a wider range of the human-likeness continuum than those other instruments. Future development and use of the scale may improve our understanding of the concept anthropomorphism.

“I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted.”

Alan Turing (1912 – 1954)

5

General discussion

The work in this dissertation was part of a long-term research project for promoting energy conservation, and it was aimed at designing persuasive interventions to stimulate sustainable behavior. Most, if not all, energy-related behavior occurs in interactions between people and their technological systems, so it is in these interactions where the crucial behavioral decisions are made. Persuasive interventions aimed at reducing energy use could thus be most effective within those interactions. The current research was aimed at utilizing artificial agents for effectively influencing energy conservation behavior. I argued that perceived human-likeness (i.e. anthropomorphism) of those agents could influence their persuasiveness. Therefore, perceived human-likeness of artificial agents could make them more effective in influencing sustainable behavior.

The main research question in this dissertation was: What is the relation between anthropomorphism of artificial agents and their persuasive effects? I examined how technological and psychological determinants of anthropomorphism could influence the effectiveness of artificial agents designed to persuade their users to save energy. In particular, I investigated effects of social exclusion as a psychological determinant of anthropomorphism, and consistency of social cues as a technological determinant of anthropomorphism, on the persuasiveness of artificial agents.

In Chapter 2, effects of social exclusion on susceptibility to persuasion by artificial agents were investigated by testing people's susceptibility to social feedback coming from an artificial agent after being socially included or excluded. I argued that being socially excluded would make people perceive artificial agents as more human-like and be more persuaded by them. The main hypothesis was that socially excluded people would become more susceptible to social feedback coming from an artificial agent. This hypothesis was tested in two studies.

Study 1 investigated effects of social exclusion on susceptibility to persuasion by a female artificial agent in a simulated washing machine task. In Study 2, this design was replicated and extended with a male artificial agent. Findings of both studies supported the main hypothesis. That is, socially excluded people adapted their behavior to a greater extent in response to the artificial agents' social feedback than socially included people in both studies. Additionally, an effect of participant gender was found. More specifically, female participants were more susceptible to the social feedback than male participants. Results did not support the expectation that social exclusion would influence measures of anthropomorphism.

In Chapter 3, effects of consistency of social cues on susceptibility to persuasion by artificial agents were investigated by testing people's responses to artificial agents that displayed two social cues that were either consistent or inconsistent with each other. I argued that perceived inconsistencies in artificial agents' social cues would make people perceive those agents as less human-like and be less persuaded by them. The hypotheses were that consistency of an artificial agent's social cues would increase people's recognition and recall of emotions conveyed by those agents and people's susceptibility to social feedback by the agents. These hypothesis were tested in two studies.

Study 3 investigated effects of consistency of artificial agents' social cues on the agents' perceived human-likeness and people's recognition of the agents' emotions. Study 4 investigated effects of consistency of an artificial agent's social cues on people's recall of the agent's social feedback and its persuasiveness in a simulated thermostat task. Findings of both studies supported the hypotheses. That is, consistent social cues were recognized faster than inconsistent ones, and, more importantly, people adapted their behavior to a greater extent in response to an artificial agent that displayed consistent social cues than to one that displayed inconsistent social cues. As in Chapter 2, no support was found for the expectation that a determinant of anthropomorphism would increase the agents' perceived human-likeness. That is, no effects of consistency of social cues were found on measures of anthropomorphism. For this reason, I will reconsider the role of anthropomorphism and its relation with persuasion in the next section.

5.1 The role of anthropomorphism

Literature on social connectedness (e.g., Pickett et al., 2004; Epley, Akalis, et al., 2008; Eyssel & Reich, 2013) would predict an increase in people's anthropomorphic perceptions of artificial agents after being socially excluded. Similarly, literature on social cues (e.g., Nass et al., 1994; Duffy, 2003; Vossen et al., 2010) would predict an increase in people's anthropomorphic perceptions of artificial agents when they show consistent social cues. The findings presented in this dissertation did not support these predictions. Were the predictions based on existing literature false? Or could there be other explanations for not finding the expected effects of social exclusion and consistency of social cues on measurements of anthropomorphism? In the next sections, I will discuss several possible explanations.

5.1.1 A different mediating process

One possible explanation for not finding the expected effects on measurements of anthropomorphism could be that anthropomorphism is not the mediating process. I argued that self-esteem might explain the effect of social exclusion on susceptibility to persuasion by artificial agents, because it was shown to be related to susceptibility for social influence (e.g., Janis, 1954; Berkowitz & Lundy, 1957). This could explain why effects of social exclusion were found on persuasion, but not on measurements of anthropomorphism.

Earlier research has shown effects of social exclusion on people's reported levels of self-esteem (Zadro et al., 2004). In that study, self-esteem was measured by asking participants how they felt about themselves *while being engaged in the game*. This does not necessarily imply that participants also experienced lower levels of self-esteem *after the game*. It seems plausible that being socially excluded during the game made participants (unconsciously) more attentive to social cues after the game. Consequently, this attentiveness to social cues could have made them more susceptible to social feedback, without influencing their explicit reports of anthropomorphism. This explanation corresponds with the assumption that anthropomorphism is an unconscious process. This unconsciousness was shown by Nass and Moon (2000), who reported that people responded socially to technology on a behavioral level, but when explicitly asked about these social responses, they stated that they would never show such behavior.

An explanation for not finding effects of consistency of social cues on measurements of anthropomorphism could be that artificial agents that displayed inconsistent cues were perceived as less credible. Source credibility is commonly used to explain people's acceptance of persuasive messages (e.g., Sternthal, Dholakia, & Leavitt, 1978; Pornpitakpan, 2004). This might explain why effects of consistency of social cues were found on persuasion, but not on measurements of anthropomorphism. However, anthropomorphism was shown to be related to credibility (e.g., Nowak, 2004; Nowak & Rauh, 2005), and therefore these concepts could equally well explain the relation between consistency of an artificial agent's social cues and its persuasiveness. Future research could more closely investigate the relation between credibility and anthropomorphism.

5.1.2 Effects too small to detect

Another possible explanation for not finding the expected effects on measurements of anthropomorphism could be that those effects were too small to detect. The experimental manipulations were designed to only influence one particular element of people's interactions with the artificial agents. Consequently, those manipulations were quite subtle, which could be the reason for not finding effects on people's explicit reports of anthropomorphism. If the manipulations were indeed too subtle to find effects on measurements of anthropomorphism, the question remains why participants did adapt their behavior to a certain extent.

In Chapter 4, I argued that available measuring instruments all focus on a specific subset of the human-likeness continuum. These subsets appeared to be in the middle and high regions of this continuum, so those measuring instruments are only able to measure medium to high levels of anthropomorphism. This could also explain the low averages found on those instruments in the experiments in this dissertation (which were around 2.0 to 3.5 on a 7-point scale).

Another reason for these low averages could be that the artificial agents were programmed to respond every time a participant changed a setting in the task. These programmed responses could have made those agents be perceived as low in their human-likeness, leading to low averages on the measurements of anthropomorphism. This poses the question to what extent the available measuring instruments are suitable for measuring low levels of anthropomorphism. In the next section, I will discuss this and further issues with those measuring instruments.

5.1.3 Issues with available measuring instruments

I argue that the most likely explanation for not finding the expected effects on measurements of anthropomorphism is related to the measurement itself. Anthropomorphism is a complex phenomenon, involving unconscious attributions of human-like characteristics to non-humans. These attributions may lead to outcomes in the form of affective states, behavioral changes, or thoughts about the anthropomorphized non-human. Such a complex psychological phenomenon might be difficult to quantify (for a discussion on the issue of quantifying psychological concepts, see e.g., Michell, 1997; Wright, 1999; Michell, 1999).

Because of its automatic components and its –at least partially– unconscious nature, anthropomorphism should not be measured by only asking whether a technology has characteristics like free will or consciousness, items that are high on the human-likeness continuum and difficult to attribute to technology. Instead, measuring instruments should also include human-like characteristics that are low on the human-likeness continuum, because they are more easily ascribed to that technology. As I already mentioned in the previous section, available measuring instruments of anthropomorphism only measure specific subsets of the continuum. Those instruments thus insufficiently measure the concept, which could explain why no effects on measurements of anthropomorphism were found. To potentially overcome this problem, a new measuring instrument was developed in Chapter 4: the Predisposition to Anthropomorphize Scale.

5.2 Predispositions to anthropomorphize

In Chapter 4, I argued that there is no convincing evidence for anthropomorphism being a multidimensional construct, and that it should therefore be viewed as a one-dimensional one. The Predisposition to Anthropomorphize Scale was designed to map various anthropomorphic characteristics on such a one-dimensional scale. The main hypotheses were that all human-like characteristics could be successfully mapped onto a one-dimensional scale and that they would be ordered in a way that is similar for all individuals in their encounter with different types of agents. These hypotheses, along with the scale's psychometric properties (i.e. its reliability and validity), were investigated in three studies.

In Study 5, a first version of the Predisposition to Anthropomorphize Scale was developed and tested on its construct validity, which is the degree to which the scale measures anthropomorphism. Construct validity was tested by comparing the estimated difficulty to attribute a specific item to a robot with its perceived level of human nature and human uniqueness. In Study 6, the Predisposition to Anthropomorphize Scale was tested on its convergent validity, which is the degree to which the scale relates to existing measuring instruments of anthropomorphism. Convergent validity was tested by comparing the Predisposition to Anthropomorphize Scale with the Godspeed- and Waytz-instruments. In Study 7, the Predisposition to Anthropomorphize Scale was tested on its sensitivity for differentiating between different agent types, which was tested by comparing the perceived human-likeness of humans, robots, computers, and algorithms.

Findings of all three studies confirmed the hypotheses. The Predisposition to Anthropomorphize Scale showed an invariant ordering of human-like characteristics on a single dimension, and significant correlations were found between the scale and existing measuring instruments of anthropomorphism. The scale also appeared to measure a wider range of the human-likeness continuum than existing measuring instruments. Finally, the Predisposition to Anthropomorphize Scale successfully differentiated between different types of agents. The scale could be used in future studies to collect reliable measurements of anthropomorphism that cover a wide range of the anthropomorphism dimension, and to be able to compare research findings.

The Predisposition to Anthropomorphize Scale contributes to the current dissertation by measuring anthropomorphism in a more reliable and objective way than other available measuring instruments. The scale is reliable because it corresponds sufficiently with concepts related to anthropomorphism (i.e. human nature and human uniqueness) and with other available measuring instruments. The scale is objective because human-like characteristics were ordered in a way that was similar for all individuals in their encounter with different types of agents. The Predisposition to Anthropomorphize Scale also demonstrated shortcomings of available measuring instruments, which may explain why the expected effects on measurements of anthropomorphism were not found in Chapters 2 and 3.

5.3 Contributions of the current work

The work presented in this dissertation contributes to the field of human-computer interaction by being –to my knowledge– the first that investigated effects of psychological and technological determinants of anthropomorphism on persuasion by artificial agents. Earlier work has investigated links between social exclusion and anthropomorphism (e.g., Epley et al., 2007; Epley, Waytz, et al., 2008; Eyssel & Reich, 2013), social cues and social responses to computers (e.g., Nass et al., 1993; Nass & Moon, 2000; Duffy, 2003), and anthropomorphism and persuasion (e.g., Zanbaka, Goolkasian, & Hodges, 2006; Nan, Anghelcev, Myers, Sar, & Faber, 2006; Yoo & Gretzel, 2011). The innovative contribution of the current work is that it investigated direct links between social exclusion and persuasion and between consistency of social cues and persuasion at the behavioral level. Additionally, a new measuring instrument for anthropomorphism was developed that measures the concept in a more reliable way than the available measuring instruments.

The general aim of Chapter 2 was to investigate the link between social exclusion and persuasion by an artificial agent. The relation between social exclusion and anthropomorphism as well as the one between anthropomorphism and persuasion were already demonstrated in previous research. Results of Chapter 2 contribute to the field of human-computer interaction by demonstrating that the psychological state of a user of PT may influence his/her responses to that technology, and could thereby influence the technology's persuasive effectiveness.

The general aim of Chapter 3 was to investigate the link between consistency of different social cues and persuasion by an artificial agent. When artificial agents are designed with multiple social cues, these cues may be perceived as inconsistent, which could make the agent being perceived as less human-like. Results of Chapter 3 contribute to the field of human-computer interaction by demonstrating that perceived inconsistencies of social cues in PT could influence a user's responses to that technology and could thereby also influence its persuasive effectiveness.

Unlike other researchers who described anthropomorphism as a multidimensional construct (e.g., Złotowski et al., 2014), I argued that all human-like characteristics that can be attributed to non-humans are invariantly ordered on a single dimension. This unidimensionality could simplify the measurement of anthropomorphism, and contribute to our understanding of the concept. Results of Chapter 4 contribute to the field of human-computer interaction by demonstrating that anthropomorphism can successfully be measured on one dimension. Moreover, differences between the Predisposition to Anthropomorphize Scale and two available measuring instruments of anthropomorphism (i.e. the Waytz- and Godspeed-instruments) were revealed. Those two measuring instruments appeared to measure only a subset of the human-likeness continuum, which makes them unsuitable for comparing different types of agents, especially when the perceived human-likeness of those agents is above or below that specific subset. Research in human-computer-interaction and human-robot-interaction could benefit from the Predisposition to Anthropomorphize Scale because it provides reliable and comparable estimations of anthropomorphism of different types of agents and robots. However, more work is needed to further develop and validate the scale and to test whether it could also be used to estimate the anthropomorphism of other types of agents than robots and computers, for example artificial agents or animals.

5.4 Recommendations of the current work

Findings in this dissertation could be used as recommendations for PT in the role of a social actor. One recommendation is to incorporate the user's psychological state in the design of PT, because this psychological state could determine a user's attentiveness to social cues, and consequently his/her susceptibility to persuasion. Earlier work already indicated the importance of individual differences in susceptibility to persuasion for the design of PT (e.g., Kaptein, Markopoulos, de Ruyter, & Aarts, 2009). More specifically, people differed in their susceptibility to different persuasion principles (as described by Cialdini, 1995), so adopting a persuasive system to these individual differences could enhance its effectiveness (Kaptein et al., 2009). Such individual differences are generally quite stable (e.g., Conley, 1985). The work in the current dissertation demonstrated that, in addition to these individual differences, variations of a person's psychological state could also influence the effectiveness of PT. That is, findings in this dissertation demonstrated that inducing such a psychological state (i.e. being socially excluded) influenced people's susceptibility to persuasion by artificial agents. More work is needed to investigate whether other psychological states could also affect people's susceptibility to persuasion by those agents. The sources of influence proposed by Epley et al. (2007) could serve as a starting point for this investigation. Several psychological states have already been investigated in the context of anthropomorphic perceptions (i.e. loneliness and need for control, see Epley, Waytz, et al., 2008), but those studies did only include effects on anthropomorphism, and not on persuasion.

Another recommendation is to incorporate users' perceptions of combined social cues in the design of PT in the role of a social actor, because the *interactions* between those cues could largely determine their perceived meaning. The importance of including social cues in the design of PT has been acknowledged before (e.g., Fogg, 2003), but the work in the current dissertation demonstrated that effects of those combined cues are not simply a sum of their individual effects. That is, perceived inconsistencies of social cues influenced people's susceptibility to persuasion by artificial agents. More work is needed to investigate the influence of different combinations of social cues on people's anthropomorphic perceptions of artificial agents and the extent to which they influence people's attitudes and, ultimately, their behavior.

5.5 Limitations and future work

While investigating the relation between anthropomorphism of artificial agents and their persuasive effects, my experiences with the available measuring instruments for anthropomorphism were quite unsatisfactory. In Chapter 4, I argued that those instruments only measured a subset of the human-likeness continuum, mainly because different researchers appeared to have different interpretations of the concept. To solve this problem, I started developing a new measuring instrument that maps various human-like characteristics on a one-dimensional scale ranging from low to high in perceived human-likeness. This instrument was developed after the empirical work on the psychological and the technological determinants was performed, and could thus not be used to investigate the hypothesized relation between anthropomorphism and persuasion.

The Predisposition to Anthropomorphize Scale was developed and tested within experimental designs that investigated people's responses to different types of robots and computers. Consequently, this instrument cannot be used in its current form to investigate people's anthropomorphic perceptions of artificial agents. Moreover, findings with the scale in the context of robots may be difficult to translate into the context of artificial agents. Future research could be designed to investigate whether the Predisposition to Anthropomorphize Scale could also be used to estimate people's anthropomorphic perceptions of artificial agents.

The work in the current dissertation demonstrated that females were more susceptible to the social feedback provided by artificial agents than males. However, this conclusion was based on the finding that females *used less energy* than males, so it could also be the case that females were more energy-friendly than males. Which of these two explanations (being more susceptible to social feedback or more energy-friendly) is most likely cannot be concluded from the work in this dissertation. Earlier research showed that females reported more environmental behaviors than males (Olli, Grendstad, & Wollebaek, 2001). In addition, women were more concerned about environmental problems and more willing to change their behavior accordingly than men (Kollmuss & Agyeman, 2002). This willingness for behavioral change could be related to susceptibility to social influence, because females were also shown to be more susceptible to persuasion than males (e.g., Eagly, 1978; Orji, 2014). Because of the salience of sustainability in today's society, these two factors (energy-use and susceptibility) are difficult to disconnect. To be able to separate those factors, future research could be designed to investigate whether females who are high or low in susceptibility to social influence differ in their sustainable behavior, and reversely, whether those who show much or little sustainable behavior differ in their susceptibility to social influence.

The employed manipulations were, for experimental reasons, rather subtle. For example, the concept of chronic loneliness was manipulated by having participants play a short simulated ball-tossing game with two computerized players. Because of this, effect sizes were expected (and found) to be quite small. The experimental designs also did not enable me to investigate whether people who are truly lonely respond differently to persuasion by artificial agents than people who are not. Future research could be designed to investigate whether such effects are stronger than the ones reported in this dissertation. If so, this would suggest that the effects found in controlled lab environments are translatable to dynamic home environments, which could largely influence the practical potential of artificial agents as PT.

The studies presented in the current dissertation had samples of around 20-35 participants per between-subjects condition. While this is quite a common practice in psychological research, there has been a lively debate on the influence of small sample sizes on research findings (e.g., Ioannidis, 2008; Simmons, Nelson, & Simonsohn, 2011; Nelson, Simmons, & Simonsohn, 2012; Simmons, Nelson, & Simonsohn, 2013; Lakens & Evers, 2014), especially because the statistical power of psychological studies is argued to be rather low (e.g., Cohen, 1962; Rossi, 1990; Cohen, 1992). These issues could raise questions about the reproducability of the presented findings. However, the effects of both social exclusion and consistency of social cues were replicated within this dissertation, which greatly increases the probability of those effects being true effects (for overviews of the probability of replicating significant findings, see e.g., Cumming, 2008; Miller, 2009).

All studies in this dissertation were performed on a computer (either online or in controlled laboratory environments). Studies that are performed on computers are shown to be valid for testing psychological processes (e.g., Anderson, Lindsay, & Bushman, 1999; Krantz & Dalal, 2000), but it should also be acknowledged that these controlled methods have little ecological validity. In order to be able to predict how people would respond to artificial agents that provide energy-related feedback in their homes, it is important to increase this ecological validity (see e.g., Ruijten, de Kort, & Kosnar, 2012). One way of increasing the ecological validity is to perform field studies. Such studies however do not have much experimental control, because the home environment is very dynamic. Another way of increasing ecological validity, without compromising experimental control, is using immersive virtual reality (see e.g., Loomis, Blascovich, & Beall, 1999). In order to increase the predictive value of the presented studies, future research could be designed to use such immersive environments.

In addition to the ecological validity of the studies presented in this dissertation, care needs to be given to the *duration* of people's interactions with artificial agents as well. The work in this dissertation investigated short-term effects of determinants of anthropomorphism on persuasion by artificial agents. More specifically, participants briefly interacted with artificial agents, after which the extent to which they adapted their behavior according to the agents' social feedback was measured. The importance of understanding long-term effects of PT has been stressed before (e.g., IJsselsteijn, de Kort, Midden, Eggen, & van den Hoven, 2006; Reitberger, Meschtscherjakov, Tscheligi, de Ruyter, & Ham, 2010), but they are still hardly investigated. Long-term effects of providing persuasive messages or electronic feedback have for example been investigated in a seven-month field study on home energy use (Kluckner, Weiss, Schrammel, & Tscheligi, 2013) and a two week text-messaging intervention on snacking behavior (Kaptein, De Ruyter, Markopoulos, & Aarts, 2012). These studies however used ambient feedback and text messages instead of interactive artificial agents. To increase the persuasiveness of PT in the form of artificial agents, future studies could be designed to investigate long-term effects.

5.6 Ethical considerations

In his description of PT, Fogg (1999) acknowledged that it raises several ethical concerns by emphasizing that it should not coerce or deceive its users while fulfilling its persuasive goals. Most PT is designed to make people save energy or live healthier or safer lives, of which most would agree are beneficial for everyone. Nevertheless, PT does influence its users' behavior, and even though designers may have good intentions, ethical considerations should be taken into account. In Chapter 2, I investigated the relation between a psychological state (feeling socially excluded) and persuasion. Participants were induced to feel socially excluded, after which they were found to be more susceptible to persuasion by an artificial agent. Inducing a psychological state in order to influence a person's behavior cannot be regarded as being very ethical. It is therefore important to consider that PT that is designed to influence people's behavior at home should not *induce* such a psychological state, but instead use sensors that can *detect* how a person is feeling, and use that information to *adapt* its persuasive strategies. An example of PT that adapts its persuasive strategies without inducing psychological states is persuasion profiling (see e.g., Kaptein & Eckles, 2010). With persuasion profiling, PT can adapt its persuasive strategies to individual users, which could increase its effectiveness without inducing psychological states.

5.7 Conclusions

The work presented in the current dissertation adds to the understanding of how psychological and technological determinants of anthropomorphism could influence persuasion by artificial agents. Further research is needed to investigate whether persuasion by artificial agents is effective in (longitudinal) field applications. Additionally, the newly developed measuring instrument for anthropomorphism could help researchers comparing their findings, because it measures the concept in a more reliable way than other existing measuring instruments. Further research is needed to investigate whether the Predisposition to Anthropomorphize Scale could also predict behavior changes.

When Alan Turing proposed the 'imitation game' (Turing, 1950) which has later come to be known as the Turing Test, he predicted that it would take approximately 50 years before a computer would be able to convince people that it is a human. In June 2014, a computer passed the proposed 30% threshold for the first time, convincing one third of its judges that it was a 13-year-old boy. Computers nowadays are not only automatically perceived as human-like, they actually convince people that they are, in fact, human. In the next 50 years, such smart computers could be given more human-like appearances, and artificial agents may become more human-like, slowly blurring the boundaries between people and technology. Although it is hard to predict whether these boundaries will ever fully disappear, I hope the work in this dissertation contributes to our understanding of social interactions between humans and technology.

References

Abrahamse, W., Steg, L., Vlek, C., & Rothengatter, T. (2007). The effect of tailored information, goal setting, and tailored feedback on household energy use, energy-related behaviors, and behavioral antecedents. *Journal of Environmental Psychology*, *27*(4), 265–276.

Adams, R. B., & Kleck, R. E. (2003). Perceived gaze direction and the processing of facial displays of emotion. *Psychological Science*, *14*(6), 644–647.

Adams, R. B., & Kleck, R. E. (2005). Effects of direct and averted gaze on the perception of facially communicated emotion. *Emotion*, *5*(1), 3–11.

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nature Reviews Neuroscience*, *4*(3), 165–178.

Ahn, H.-K., Kim, H. J., & Aggarwal, P. (2014). Helping fellow beings anthropomorphized social causes and the role of anticipatory guilt. *Psychological Science*, *25*(1), 224–229.

Allen, J. P., Chango, J., Szwedo, D., Schad, M., & Marston, E. (2012). Predictors of susceptibility to peer influence regarding substance use in adolescence. *Child Development*, *83*(1), 337–350.

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: role of the sts region. *Trends in Cognitive Sciences*, *4*(7), 267–278.

Anderson, C. A., Lindsay, J. J., & Bushman, B. J. (1999). Research in the psychological laboratory: Truth or triviality? *Current Directions in Psychological Science*, *8*(1), 3–9.

Andrich, D. (1982). An extension of the rasch model for ratings providing both location and dispersion parameters. *Psychometrika*, *47*(1), 105–113.

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze.* Cambridge University Press.

Arroyo, E., Bonanni, L., & Selker, T. (2005). Waterbot: Exploring feedback and persuasive techniques at the sink. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 631–639).

Aydin, N., Fischer, P., & Frey, D. (2010). Turning to god in the face of ostracism: Effects of social exclusion on religiousness. *Personality and Social Psychology Bulletin*, *36*(6), 742–753.

Bailenson, J. N., & Yee, N. (2005). Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science*, *16*(10), 814–819.

Bargh, J. A. (1984). Automatic and conscious processing of social information. *Handbook of Social Cognition*, *3*, 1–43.

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, *1*(1), 71–81.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*(3), 497–529.

Baumeister, R. F., Twenge, J. M., & Nuss, C. K. (2002). Effects of social exclusion on cognitive processes: anticipated aloneness reduces intelligent thought. *Journal of Personality and Social Psychology*, *83*(4), 817–827.

Baylor, A. L. (2009). Promoting motivation with virtual agents and avatars: Role of visual presence and appearance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364* (1535), 3559–3565.

Bearden, W. O., Netemeyer, R. G., & Teel, J. E. (1989). Measurement of consumer susceptibility to interpersonal influence. *Journal of Consumer Research*, 473–481.

Berkowitz, L., & Lundy, R. M. (1957). Personality characteristics related to susceptibility to influence by peers or authority figures. *Journal of Personality*, *25* (3), 306–316.

Bernstein, M. J., & Claypool, H. M. (2012). Social exclusion and pain sensitivity why exclusion sometimes hurts and sometimes numbs. *Personality and Social Psychology Bulletin*, *38* (2), 185–196.

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, *13* (2), 103–124.

Bond, T. G., & Fox, C. M. (2013). *Applying the rasch model: Fundamental measurement in the human sciences.* Psychology Press.

Burger, J. M. (1992). *Desire for control: Personality, social, and clinical perspectives.* Plenum Press.

Burns, P., Lueg, C., & Berkovsky, S. (2013). Colours that move you: Persuasive ambient activity displays. In *Persuasive technology* (pp. 27–32). Springer.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42* (1), 116–131.

Carli, L. L. (2013). Gendered communication and social influence. *The SAGE Handbook of Gender and Psychology*, 199–215.

Carter-Sowell, A. R., Chen, Z., & Williams, K. D. (2008). Ostracism increases social susceptibility. *Social Influence*, *3* (3), 143–153.

Charles, E. P. (2005). The correction for attenuation due to measurement error: Clarifying concepts and creating confidence sets. *Psychological Methods*, *10*(2), 206–226.

Choi, Y. K., Miracle, G. E., & Biocca, F. (2001). The effects of anthropomorphic agents on advertising effectiveness and the mediating role of presence. *Journal of Interactive Advertising*, *2*(1), 19–32.

Christian, B. (2012). *The most human human: What artificial intelligence teaches us about being alive.* Random House LLC.

Cialdini, R. B. (1995). Principles and techniques of social influence. *Advanced Social Psychology*, *256*, 257–281.

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, *65*(3), 145–153.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Conley, J. J. (1985). Longitudinal stability of personality traits: A multitrait–multimethod–multioccasion analysis. *Journal of Personality and Social Psychology*, *49*(5), 1266–1282.

Cumming, G. (2008). Replication and p intervals: P values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science*, *3*(4), 286–300.

Dalibard, S., Magnenat-Talmann, N., Thalmann, D., et al. (2012). Anthropomorphism of artificial agents: A comparative survey of expressive design and motion of virtual characters and social robots. In *Workshop on autonomous social robots and virtual humans at the 25th annual conference on computer animation and social agents (casa 2012).*

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, *42*(3), 177–190.

Eagly, A. H. (1978). Sex differences in influenceability. *Psychological Bulletin*, *85*(1), 86–116.

Eagly, A. H., & Carli, L. L. (1981). Sex of researchers and sex-typed communications as determinants of sex differences in influenceability: A meta-analysis of social influence studies. *Psychological Bulletin*, *90*(1), 1–20.

Edwards, A. L. (1957). The social desirability variable in personality assessment and research.

Ekman, P. (1971). Universals and cultural differences in facial expressions of emotion. In *Nebraska symposium on motivation.* University of Nebraska Press.

Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, *24*(6), 581–604.

Epley, N., Akalis, S., Waytz, A., & Cacioppo, J. T. (2008). Creating social connection through inferential reproduction loneliness and perceived agency in gadgets, gods, and greyhounds. *Psychological Science*, *19*(2), 114–120.

Epley, N., Waytz, A., Akalis, S., & Cacioppo, J. T. (2008). When we need a human: Motivational determinants of anthropomorphism. *Social Cognition*, *26*(2), 143–155.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological review*, *114*(4), 864–886.

Evans, J. S. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in cognitive sciences*, *7*(10), 454–459.

Eyssel, F., Hegel, F., Horstmann, G., & Wagner, C. (2010). Anthropomorphic inferences from emotional nonverbal cues: A case study. In *Ro-man, 2010 ieee* (pp. 646–651).

Eyssel, F., Kuchenbrandt, D., & Bobinger, S. (2011). Effects of anticipated human-robot interaction and predictability of robot behavior on perceptions of anthropomorphism. In *Proceedings of the 6th international conference on human-robot interaction* (pp. 61–68).

Eyssel, F., & Reich, N. (2013). Loneliness makes the heart grow fonder (of robots): On the effects of loneliness on psychological anthropomorphism. In *8th acm/ieee international conference on human-robot interaction (hri), 2013* (pp. 121–122). IEEE.

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research*, 303–315.

Fogg, B. J. (1999). Persuasive technologies. *Communications of the ACM*, *42*(5), 26–29.

Fogg, B. J. (2003). *Persuasive technology: Using computers to change what we think and do.* Morgan Kaufmann.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, *42*(3), 143–166.

Frith, C. D., & Frith, U. (2007). Social cognition in humans. *Current Biology*, *17*(16), 724–732.

Frith, U., & Frith, C. D. (2010). The social brain: Allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1537), 165–176.

Guadagno, R. E., Blascovich, J., Bailenson, J. N., & Mccall, C. (2007). Virtual humans and persuasion: The effects of agency and behavioral realism. *Media Psychology*, *10*(1), 1–22.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, *3*(4), 367–388.

Guthrie, S. (1993). *Faces in the clouds: A new theory of religion.* Oxford University Press.

Haans, A., Kaiser, F. G., Bouwhuis, D. G., & IJsselsteijn, W. A. (2012). Individual differences in the rubber-hand illusion: Predicting self-reports of people's personal experiences. *Acta Psychologica*, *141*(2), 169–177.

Ham, J., Midden, C. J. H., Maan, S. J., & Merkus, B. (2009). Persuasive lighting: The influence of feedback through lighting on energy conservation behavior. *Experiencing Light 2009: International Conference on the Effects of Light on Wellbeing*, 122–128.

Harmon-Jones, E., Peterson, C. K., & Harris, C. R. (2009). Jealousy: Novel methods and neural correlates. *Emotion*, *9*(1), 113–117.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, *108*(3), 356–388.

Haslam, N., Bain, P., Douge, L., Lee, M., & Bastian, B. (2005). More human than you: Attributing humanness to self and others. *Journal of Personality and Social Psychology*, *89*(6), 937–950.

Haslam, N., Loughnan, S., Kashima, Y., & Bain, P. (2008). Attributing and denying humanness to others. *European Review of Social Psychology*, *19*(1), 55–85.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, *57*(2), 243–259.

Herring, H., & Roy, R. (2007). Technological innovation, energy efficient design and the rebound effect. *Technovation*, *27*(4), 194–203.

Hertwich, E. G. (2005). Consumption and the rebound effect: An industrial ecology perspective. *Journal of Industrial Ecology*, *9*(1-2), 85–98.

Hume, D. (1889). *The natural history of religion.* London: A. and H. Bradlaugh Bonner.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings.* Sage.

IJsselsteijn, W., de Kort, Y. A. W., Midden, C. J. H., Eggen, B., & van den Hoven, E. (2006). Persuasive technology for human well-being: Setting the scene. In *Persuasive technology* (pp. 1–5). Springer.

Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology*, *19*(5), 640–648.

Jaffe, A. B., & Stavins, R. N. (1994). The energy paradox and the diffusion of conservation technology. *Resource and Energy Economics*, *16*(2), 91–122.

Janis, I. L. (1954). Personality correlates of susceptibility to persuasion. *Journal of Personality*, *22*(4), 504–518.

Johnson, R. D., Veltri, N. F., & Hornik, S. (2008). Attributions of responsibility toward computing technology: The role of interface social cues and user gender. *International Journal of Human–Computer Interaction*, *24*(6), 595–612.

Jung, Y., & Lee, K. M. (2004). Effects of physical embodiment on social presence of social robots. *Proceedings of PRESENCE*, 80–87.

Kappel, K., & Grechenig, T. (2009). Show-me: Water consumption at a glance to promote water conservation in the shower. In *Proceedings of the 4th international conference on persuasive technology* (pp. 26:1–26:6).

Kaptein, M., De Ruyter, B., Markopoulos, P., & Aarts, E. (2012). Adaptive persuasive systems: A study of tailored persuasive text messages to reduce snacking. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *2*(2), 10:1–10:25.

Kaptein, M., & Eckles, D. (2010). Selecting effective means to any end: Futures and ethics of persuasion profiling. In *Persuasive technology* (pp. 82–93). Springer.

Kaptein, M., Markopoulos, P., de Ruyter, B., & Aarts, E. (2009). Can you be persuaded? individual differences in susceptibility to persuasion. In *Human-computer interaction–interact 2009* (pp. 115–118). Springer.

Kaptein, M., & van Halteren, A. (2013). Adaptive persuasive messaging to increase service retention: Using persuasion profiles to increase the effectiveness of email reminders. *Personal and Ubiquitous Computing*, *17*(6), 1173–1185.

Katagiri, Y., Nass, C., & Takeuchi, Y. (2001). Cross-cultural studies of the computers are social actors paradigm: The case of reciprocity. *Usability Evaluation and Interface Design: Cognitive Engineering, Intelligent Agents, and Virtual Reality*, 1558–1562.

Kennedy, J. S. (1992). *The new anthropomorphism.* Cambridge University Press.

Khazzoom, J. D. (1980). Economic implications of mandated efficiency in standards for household appliances. *The Energy Journal*, *1*(4), 21–40.

Kidd, C. D., & Breazeal, C. (2004). Effect of a robot on user perceptions. In *Intelligent robots and systems, 2004.(iros 2004). proceedings. 2004 ieee/rsj international conference on* (Vol. 4, pp. 3559–3564).

Kim, C., & Baylor, A. L. (2008). A virtual change agent: Motivating pre-service teachers to integrate technology in their future classrooms. *Journal of Educational Technology & Society*, *11*(2), 309–321.

Kim, T., Hong, H., & Magerko, B. (2010). Designing for persuasion: Toward ambient eco-visualization for awareness. In *Persuasive technology* (pp. 106–116). Springer.

Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin*, *100*(1), 78–100.

Kluckner, P. M., Weiss, A., Schrammel, J., & Tscheligi, M. (2013). Exploring persuasion in the home: Results of a long-term study on energy consumption behavior. In *Ambient intelligence* (pp. 150–165). Springer.

Knapp, M., Hall, J., & Horgan, T. (2013). *Nonverbal communication in human interaction.* Cengage Learning.

Kollmuss, A., & Agyeman, J. (2002). Mind the gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? *Environmental Education Research*, *8*(3), 239–260.

Krantz, J. H., & Dalal, R. (2000). Validity of web-based psychological research. In M. Birnbaum (Ed.), *Psychological experiments on the internet.* New York: Academic Press.

Lakens, D., & Evers, E. R. K. (2014). Sailing from the seas of chaos into the corridor of stability practical recommendations to increase the informational value of studies. *Perspectives on Psychological Science*, *9*(3), 278–292.

Langton, S. R. H., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? cues to the direction of social attention. *Trends in Cognitive Sciences*, *4*(2), 50–59.

Leary, M. R., Kelly, K. M., Cottrell, C. A., & Schreindorfer, L. S. (2007). Individual differences in the need to belong: Mapping the nomological network. *Unpublished manuscript*.

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *The Journal of Economic Perspectives*, 153–174.

Linacre, J. M. (2001). Category, step and threshold: Definitions & disordering. *Rasch Measurement Transactions*, *15*(1).

Linacre, J. M. (2003). Data variance: Explained, modeled and empirical. *Rasch Measurement Transactions*, *17*(3), 942–943.

Lockton, D., Harrison, D., & Stanton, N. (2008). Making the user more efficient: Design for sustainable behaviour. *International Journal of Sustainable Engineering*, *1*(1), 3–8.

Loomis, J. M., Blascovich, J. J., & Beall, A. C. (1999). Immersive virtual environment technology as a basic research tool in psychology. *Behavior Research Methods, Instruments, & Computers*, *31*(4), 557–564.

Lutzenhiser, L. (1993). Social and behavioral aspects of energy use. *Annual Review of Energy and the Environment*, *18*(1), 247–289.

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle.* (Vol. 1). MIT Press.

McCalley, T. L., Kaiser, F., Midden, C. J. H., Keser, M., & Teunissen, M. (2006). Persuasive appliances: Goal priming and behavioral response to product-integrated energy feedback. In *Persuasive technology* (pp. 45–49). Springer.

McCalley, T. L., & Midden, C. J. H. (2002). Energy conservation through product-integrated feedback: The roles of goal-setting and social orientation. *Journal of economic psychology*, *23*(5), 589–603.

Meijer, R. R., & Sijtsma, K. (2001). Methodology review: Evaluating person fit. *Applied Psychological Measurement*, *25*(2), 107–135.

Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, *88*(3), 355–383.

Michell, J. (1999). *Measurement in psychology.* Cambridge University Press.

Midden, C. J. H., & Ham, J. (2008). The persuasive effects of positive and negative social feedback from an embodied agent on energy conservation behavior. In *Aisb 2008 convention communication, interaction and social intelligence* (Vol. 1, pp. 9–13).

Midden, C. J. H., & Ham, J. (2009). Using negative and positive social feedback from a robotic agent to save energy. In *Proceedings of the 4th international conference on persuasive technology* (pp. 12:1–12:6). ACM.

Midden, C. J. H., & Ham, J. (2012). The illusion of agency: The influence of the agency of an artificial agent on its persuasive power. In *Persuasive technology: Design for health and safety* (pp. 90–99). Springer.

Midden, C. J. H., & Ham, J. (2013). Persuasive technology to promote pro-environmental behaviour. In L. Steg, A. van den Berg, & J. de Groot (Eds.), *Environmental psychology an introduction.* BPS Blackwell.

Midden, C. J. H., Kaiser, F. G., & McCalley, T. L. (2007). Technology's four roles in understanding individuals' conservation of natural resources. *Journal of Social Issues*, *63*(1), 155–174.

Midden, C. J. H., McCalley, T. L., Ham, J., & Zaalberg, R. (2008). Using persuasive technology to encourage sustainable behavior. In *6th international conference on pervasive computing.*

Miller, J. (2009). What is the probability of replicating a statistically significant effect? *Psychonomic Bulletin & Review*, *16*(4), 617–640.

Nan, X., Anghelcev, G., Myers, J. R., Sar, S., & Faber, R. (2006). What if a web site can talk? exploring the persuasive effects of web-based anthropomorphic agents. *Journalism & Mass Communication Quarterly*, *83*(3), 615–631.

Nass, C., & Brave, S. (2005). *Wired for speech: How voice activates and advances the human-computer relationship.* MIT Press.

Nass, C., & Gong, L. (2000). Speech interfaces from an evolutionary perspective. *Communications of the ACM*, *43*(9), 36–43.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, *43*(2), 223–239.

Nass, C., Steuer, J., Tauber, E., & Reeder, H. (1993). Anthropomorphism, agency, and ethopoeia: Computers as social actors. In *Interact'93 and chi'93 conference companion on human factors in computing systems* (pp. 111–112). ACM.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 72–78). ACM.

Nelson, L. D., Simmons, J. P., & Simonsohn, U. (2012). Let's publish fewer papers. *Psychological Inquiry*, *23*(3), 291–293.

Nowak, K. L. (2004). The influence of anthropomorphism and agency on social judgment in virtual environments. *Journal of Computer-Mediated Communication*, *9*(2).

Nowak, K. L., & Rauh, C. (2005). The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction. *Journal of Computer-Mediated Communication*, *11*(1), 153–178.

Olli, E., Grendstad, G., & Wollebaek, D. (2001). Correlates of environmental behaviors bringing back social context. *Environment and Behavior*, *33*(2), 181–208.

Orji, R. (2014). Exploring the persuasiveness of behavior change support strategies and possible gender differences. In *Proceedings of the second international workshop on behavior change support systems (bcss2014), padova, italy.*

Payne, B. K., & Gawronski, B. (2010). A history of implicit social cognition: Where is it coming from? where is it now? where is it going? *Handbook of Implicit Social Cognition: Measurement, Theory, and Applications*, 1–15.

Perrett, D., Smith, P., Potter, D., Mistlin, A., Head, A., Milner, A., & Jeeves, M. (1983). Neurones responsive to faces in the temporal cortex: Studies of functional organization, sensitivity to identity and relation to perception. *Human Neurobiology*, *3*(4), 197–208.

Petkov, P., Köbler, F., Foth, M., & Krcmar, H. (2011). Motivating domestic energy conservation through comparative, community-based feedback in mobile and social media. In *Proceedings of the 5th international conference on communities and technologies* (pp. 21–30).

Pickett, C. L., Gardner, W. L., & Knowles, M. (2004). Getting a cue: The need to belong and enhanced sensitivity to social cues. *Personality and Social Psychology Bulletin*, *30*(9), 1095–1107.

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology*, *34*(2), 243–281.

Powers, A., & Kiesler, S. (2006). The advisor robot: Tracing people's mental model from a robot's physical attributes. In *Proceedings of the 1st acm sigchi/sigart conference on human-robot interaction* (pp. 218–225). ACM.

Qiu, L., & Benbasat, I. (2009). Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of Management Information Systems*, *25*(4), 145–182.

Reeves, B., & Nass, C. (1996). *How people treat computers, television, and new media like real people and places.* CSLI Publications and Cambridge University Press.

Reitberger, W., Meschtscherjakov, A., Tscheligi, M., de Ruyter, B., & Ham, J. (2010). Measuring (ambient) persuasive technologies.

Roland, P. E. (1993). *Brain activation.* Wiley-Liss.

Rossi, J. S. (1990). Statistical power of psychological research: What have we gained in 20 years? *Journal of Consulting and Clinical psychology*, *58* (5), 646–656.

Roubroeks, M., Ham, J., & Midden, C. J. H. (2011). *Investigating the media equation hypothesis: Do we really see computer agents as human-like?* (Unpublished manuscript)

Ruijten, P. A. M., de Kort, Y. A. W., & Kosnar, P. (2012). Bridging the gap between the home and the lab: A qualitative study of acceptance of an avatar feedback system. In *Persuasive technology. design for health and safety* (pp. 251–255). Springer.

Sallnäs, E.-L. (2005). Effects of communication mode on social presence, virtual presence, and performance in collaborative virtual environments. *Presence: Teleoperators and Virtual Environments*, *14* (4), 434–449.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. detection, search, and attention. *Psychological Review*, *84* (1), 1–66.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, *18* (5), 429–434.

Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology*, *52* (6), 1061–1086.

Shaw, F. (1991). Fits about "misfit". *Rasch Measurement Transactions*, *5* (1).

Shibata, T., Wada, K., Ikeda, Y., & Sabanovic, S. (2009). Cross-cultural studies on subjective evaluation of a seal robot. *Advanced Robotics*, *23* (4), 443–458.

Siegel, M., Breazeal, C., & Norton, M. I. (2009). Persuasive robotics: The influence of robot gender on human behavior. In *Intelligent robots* *and systems, 2009. iros 2009. ieee/rsj international conference on* (pp. 2563–2568).

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2013). Life after p-hacking. In *Meeting of the society for personality and social psychology, new orleans, la* (pp. 17–19).

Smith Jr, E. V. (2002). Understanding rasch measurement: Detecting and evaluating the impact of multidimenstionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, *3*(2), 205–231.

Sproull, L., Subramani, M., Kiesler, S., Walker, J. H., & Waters, K. (1996). When the interface is a face. *Human-Computer Interaction*, *11*(2), 97–124.

Steg, L. (2008). Promoting household energy conservation. *Energy Policy*, *36*(12), 4449–4453.

Stern, P. C. (1992). What psychology knows about energy conservation. *American Psychologist*, *47*(10), 1224–1232.

Sternthal, B., Dholakia, R., & Leavitt, C. (1978). The persuasive effect of source credibility: Tests of cognitive response. *Journal of Consumer Research*, *4*(4), 252–260.

Stiff, J. B., & Mongeau, P. A. (2003). *Persuasive communication.* Guilford Press.

Surakka, V., & Vanhala, T. (2011). Emotions in human-computer interaction. *Face-to-face Communication over the Internet: Emotions in a Web of Culture, Language, and Technology*, 213–236.

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

Twenge, J. M., Baumeister, R. F., DeWall, C. N., Ciarocco, N. J., & Bartels, J. M. (2007). Social exclusion decreases prosocial behavior. *Journal of Personality and Social Psychology*, *92*(1), 56–66.

Twenge, J. M., Baumeister, R. F., Tice, D. M., & Stucke, T. S. (2001). If you can't join them, beat them: Effects of social exclusion on aggressive behavior. *Journal of Personality and Social Psychology*, *81*(6), 1058–1069.

van Beest, I., & Williams, K. D. (2006). When inclusion costs and ostracism pays, ostracism still hurts. *Journal of Personality and Social Psychology*, *91*(5), 918–928.

Vossen, S., Ham, J., & Midden, C. J. H. (2009). Social influence of a persuasive agent: The role of agent embodiment and evaluative feedback. In *Proceedings of the 4th international conference on persuasive technology* (pp. 46:1–46:7). ACM.

Vossen, S., Ham, J., & Midden, C. J. H. (2010). What makes social feedback from a robot work? disentangling the effect of speech, physical appearance and evaluation. In *Persuasive technology* (pp. 52–57). Springer.

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? the stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, *5*(3), 219–232.

Waytz, A., Epley, N., & Cacioppo, J. T. (2010). Social cognition unbound insights into anthropomorphism and dehumanization. *Current Directions in Psychological Science*, *19*(1), 58–62.

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, 113–117.

Waytz, A., Morewedge, C. K., Epley, N., Monteleone, G., Gao, J.-H., & Cacioppo, J. T. (2010). Making sense by making sentient: Effectance motivation increases anthropomorphism. *Journal of Personality and Social Psychology*, *99*(3), 410–435.

Webler, T., & Tuler, S. P. (2010). Getting the engineering right is not always enough: Researching the human dimensions of the new energy technologies. *Energy Policy*, *38*(6), 2690–2691.

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology*, *67*(6), 1049–1062.

Whelan, R. (2010). Effective analysis of reaction time data. *The Psychological Record*, *58*(3), 475–482.

Williams, K. D., Cheung, C. K. T., & Choi, W. (2000). Cyberostracism: Effects of being ignored over the internet. *Journal of Personality and Social Psychology*, *79*(5), 748–762.

Williams, K. D., & Jarvis, B. (2006). Cyberball: A program for use in research on interpersonal ostracism and acceptance. *Behavior Research Methods*, *38*(1), 174–180.

Williams, K. D., & Sommer, K. L. (1997). Social ostracism by coworkers: Does rejection lead to loafing or compensation? *Personality and Social Psychology Bulletin*, *23*(7), 693–706.

Wood, G., & Newborough, M. (2003). Dynamic energy-consumption indicators for domestic appliances: Environment, behaviour and design. *Energy and Buildings*, *35*(8), 821–841.

Wright, B. D. (1977). Solving measurement problems with the rasch model. *Journal of Educational Measurement*, *14*(2), 97–116.

Wright, B. D. (1999). Fundamental measurement for psychology. In *The new rules of measurement: What every psychologist and educator should know* (pp. 65–104). Psychology Press.

Wright, B. D., Linacre, J. M., Gustafson, J. E., & Martin-Lof, P. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, *8*(3).

Xu, R. (2003). Measuring explained variation in linear mixed effects models. *Statistics in Medicine*, *22*(22), 3527–3541.

Yoo, K.-H., & Gretzel, U. (2011). Creating more credible and persuasive recommender systems: The influence of source characteristics on recommender system evaluations. In *Recommender systems handbook* (pp. 455–477). Springer.

Zaalberg, R., & Midden, C. J. H. (2010). Enhancing human responses to climate change risks through simulated flooding experiences. In *Persuasive technology* (pp. 205–210). Springer.

Zadro, L., Williams, K. D., & Richardson, R. (2004). How low can you go? ostracism by a computer is sufficient to lower self-reported levels of belonging, control, self-esteem, and meaningful existence. *Journal of Experimental Social Psychology*, *40*(4), 560–567.

Zanbaka, C., Goolkasian, P., & Hodges, L. (2006). Can a virtual cat persuade you?: The role of gender and realism in speaker persuasiveness. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 1153–1162).

Zapico, J. L., Turpeinen, M., & Brandt, N. (2009). Climate persuasive services: Changing behavior towards low-carbon lifestyles. In *Proceedings of the 4th international conference on persuasive technology* (pp. 14:1–14:8).

Zawieska, K., Duffy, B. R., & Strońska, A. (2012). Understanding anthropomorphisation in social robotics. *Pomiary, Automatyka, Robotyka*, *16*, 78–82.

Zemeckis, R. (Director). (2000). *Cast away.* Twentieth Century Fox.

Złotowski, J., Strasser, E., & Bartneck, C. (2014). Dimensions of anthropomorphism: From humanness to humanlikeness. In *Proceedings of the 2014 acm/ieee international conference on human-robot interaction* (pp. 66–73).

Summary

Responses to human-like artificial agents

The work in this dissertation was part of a long-term research project for promoting energy conservation, and it was aimed at designing persuasive interventions to stimulate sustainable behavior. Most, if not all, energy-related behavior occurs in interactions between people and their technological systems. Research has shown that people tend to respond to such interactions in similar ways as they respond to other humans, presumably because they automatically attribute human-like characteristics to smart technology. This attribution of human-like characteristics to technology and other non-human objects is defined as anthropomorphism.

The main research question in this dissertation was: What is the relation between anthropomorphism of artificial agents and their persuasive effects? I examined how technological and psychological determinants of anthropomorphism could influence the effectiveness of artificial agents designed to persuade their users to save energy. In addition, issues regarding the measurement of anthropomorphism were explored, and a new tool for reliable measurement of the concept was developed.

Chapter 2 describes the role of social connectedness as a *psychological* determinant of anthropomorphism. When people feel socially isolated, they are more likely to attribute human-like characteristics to objects. The effects of social exclusion on susceptibility to persuasion by an artificial agent were investigated in two studies in which participants performed

an energy-saving task while receiving social feedback from an artificial agent. Participants were either socially included or excluded by playing a short computerized ball tossing game. Consistent with the hypothesis, results showed an effect of social exclusion on participants' behavior in the energy-saving task.

Chapter 3 describes the role of social cues as *technological* determinants of anthropomorphism. Using multiple social cues in artificial agents could make people perceive those cues as inconsistent with each other, which might decrease the agents' persuasiveness. The effects of consistency of an artificial agent's social cues on recognition of the agent's emotions and its persuasiveness were investigated in two studies. In the first study, participants categorized artificial agents' emotional expressions, and in the second study they performed an energy-saving task while receiving social feedback from an artificial agent. Consistent with the hypotheses, consistency of artificial agents' social cues influenced recognition of its emotions and participants' energy-saving behavior.

While investigating the relation between anthropomorphism of artificial agents and their persuasiveness, the available measuring instruments for anthropomorphism were quite unsatisfactory. I argued that those instruments only measured a subset of the human-likeness continuum. Chapter 4 describes the development of a new measuring instrument based on the Rasch model that maps various human-like characteristics on a one-dimensional scale ranging from low to high in perceived human-likeness. Results of three studies showed that anthropomorphism can be measured as a range of human-like characteristics on a one-dimensional scale.

The work presented in this dissertation showed how psychological and technological determinants of anthropomorphism could influence persuasion by artificial agents. Additionally, the newly developed measuring instrument for anthropomorphism could help researchers comparing their findings, because it measures the concept on a wider range of the continuum than existing measuring instruments.

Samenvatting

Het werk in dit proefschrift was onderdeel van een onderzoeksproject voor het bevorderen van energiebesparing, en had als doel om persuasieve interventies te ontwikkelen die duurzaam gedrag stimuleren. Vrijwel al het energiegerelateerd gedrag vindt plaats in interacties tussen mensen en hun technologische systemen. Onderzoek heeft aangetoond dat men in deze interacties op een vergelijkbaare wijze reageert als in interacties met anderen, vermoedelijk omdat men automatisch menselijke eigenschappen aan slimme technologie toeschrijft. Het toeschrijven van deze eigenschappen aan technologie en andere objecten wordt gedefiniëerd als antropomorfisme.

De hoofdvraag in dit proefschrift was: Wat is de relatie tussen antropomorfisme van 'artificial agents' en hun persuasieve effecten? Ik heb onderzocht hoe technologische en psychologische determinanten van antropomorfisme de effectiviteit van 'artificial agents' die ontworpen zijn om mensen te overtuigen energie te besparen kan beïnvloeden. Daarnaast zijn kwesties omtrent het meten van antropomorfisme geëxploreerd, en is er een betrouwbaar nieuw meetinstrument voor dit concept ontwikkeld.

Hoofdstuk 2 beschrijft de rol van sociale verbondenheid als een *psychologische* determinant van antropomorfisme. Wanneer mensen zich geïsoleerd voelen hebben ze meer de neiging om menselijke eigenschappen aan objecten toe te schrijven. De effecten van sociale exclusie op de vatbaarheid voor overtuiging door een 'agent' werden onderzocht in twee studies waarin proefpersonen een energiebesparende taak op de computer deden terwijl zij sociale feedback kregen van een 'agent'. Proefpersonen werden vooraf sociaal opgenomen of uitgesloten door het spelen van een kort geautomatiseerd balspelletje. Resultaten waren in lijn met de hypothese en lieten zien dat sociale exclusie het gedrag van proefpersonen in de energie-besparende taak beïnvloedt.

Hoofdstuk 3 beschrijft de rol van sociale signalen als *technologische* determinanten van antropomorfisme. Het opnemen van meerdere sociale signalen kan ervoor zorgen dat deze worden gezien als inconsistent met elkaar, waardoor de overtuigingskracht van een ‘agent’ daalt. De effecten van consistentie van sociale signalen op de herkenning van emoties van een ‘agent’ en zijn overtuigingskracht werden onderzocht in twee studies. In de eerste studie categoriseerden proefpersonen emotionele uitdrukkingen van een ‘agent’, en in de tweede studie voerden zij een energiebesparende taak uit terwijl ze sociale feedback kregen van een ‘agent’. Resultaten waren in lijn met de hypotheses en lieten zien dat consistentie van sociale signalen zowel de herkenning van emoties als het energiezuinige gedrag van de proefpersonen versterkte.

Tijdens het onderzoek naar de relatie tussen antropomorfisme en overtuigingskracht door ‘agents’ bleken de meetinstrumenten niet optimaal, met name omdat ze slechts een beperkt deel van het antropomorfisme-continuüm konden meten. Hoofdstuk 4 beschrijft de ontwikkeling van een nieuw meetinstrument gebaseerd op het Rasch model dat verscheidene menselijke eigenschappen van laag tot hoog in menselijkheid op een eendimensionele schaal positioneert. Resultaten van drie studies lieten zien dat antropomorfisme gemeten kan worden als een reeks van menselijke eigenschappen geordernd op een eendimensionale schaal.

De bevindingen in dit proefschrift laten zien hoe psychologische en technologische determinanten van antropomorfisme de overtuigingskracht van ‘agents’ kunnen beïnvloeden. Bovendien kan het nieuw ontwikkelde meetinstrument onderzoekers helpen bij het vergelijken van onderzoeksresultaten, omdat het antropomorfisme op een betrouwbaardere manier meet dan bestaande meetinstrumenten.

List of Figures

1.1 Examples of objects that resemble human-like appearances, with (a) a road sign that provides information about a vehicle's speed with a smiley face and (b) a house that appears to have eyes. 6
1.2 The washing machine in my home which, by the design of the buttons, makes me attribute human-likeness to it. . . 16

2.1 Screenshot of the Cyberball game with the participant (Me) at the bottom and the two computerized players (Herman and Jolanda) at the top. In the screenshot, player Jolanda is throwing the ball to player Herman. . . 27
2.2 The artificial agent Femke used in Studies 1 and 2. . . . 28
2.3 Screenshot of the simulated washing machine interface. . 29
2.4 Visualization of (a) ball-percentage and (b) game-evaluation per Cyberball condition in Study 1. Whiskers represent 95% error bars. 32
2.5 Visualization of energy-conservation per (a) experimental condition and (b) participant gender in Study 1. Whiskers represent 95% confidence intervals. 33
2.6 Visualization of averaged anthropomorphism scores on the Godspeed-instrument per participant gender in Study 1. Whiskers represent 95% error bars. 35
2.7 The artificial agent Max used in Study 2. 40

2.8 Visualization of (a) ball-percentage and (b) game-evaluation per Cyberball condition in Study 2. Whiskers represent 95% error bars. 42
2.9 Visualization of energy-conservation (a) per Cyberball condition, (b) per participant gender, and (c) the interaction between Cyberball and gender-matching in Study 2. Whiskers represent 95% confidence intervals. 44
2.10 Visualization of averaged scores on the Waytz-instrument per Cyberball condition in Study 2. Whiskers represent 95% error bars. 45
2.11 Visualization of averaged scores per participant gender on (a) the Godspeed-instrument and (b) the Waytz-instrument in Study 2. Whiskers represent 95% error bars. 46

3.1 Visualization of the interaction between expression and gaze-direction for (a) the response latency, and (b) the number of errors in Study 3. Whiskers represent 95% error bars. 66
3.2 Visualization of the interaction between expression and gaze-direction for anthropomorphism ratings in Study 3. Whiskers represent 95% error bars. 67
3.3 Screenshot of the simulated thermostat interface (left) and artificial agent Kim (right) as used in Study 4. 71
3.4 Visualization of participants' recall accuracy on (a) the general feedback and (b) the agent's emotional expressions per experimental condition in Study 4. Whiskers represent 95% confidence intervals. 74
3.5 Visualization of energy-conservation per experimental condition in Study 4. Whiskers represent 95% confidence intervals. 75
3.6 Visualization of energy-conservation per participant gender in Study 4. Whiskers represent 95% confidence intervals. 76

4.1 Item-person map of Study 5, displaying the estimates of participant's predisposition to anthropomorphize and the item difficulty linked with each human-like characteristic mapped onto a single scale of equal additive units. Each number represents an item. Each dot represents a person. 98
4.2 Item invariance plots of the item difficulties of subjects with (a) even and odd identification numbers and (b) high and low predispositions to anthropomorphize in Study 5. Each number represents an item, corresponding with the numbers in Table 4.1. 100
4.3 Item invariance plots of the item difficulties of (a) Studies 5 and 6, and (b) persons in the appearance and cognition conditions in Study 6. Each number represents an item, corresponding with the numbers in Table 4.2. Red lines indicate 95% confidence intervals. 107
4.4 Items of the Predisposition to Anthropomorphize Scale, the Godspeed-instrument (items starting with a G), and the Waytz-instrument (items starting with a W) in Study 6 mapped onto a single scale. Each number represents an item or a Rasch-Andrich threshold (see Andrich, 1982 for details). Each dot represents a person. 110
4.5 Examples of pictures used in the (a) computer, (b) robot, (c) human, and (d) algorithm conditions. 114
4.6 Item invariance plot of the item difficulties in Studies 5 and 7. Each number represents an item. Red lines indicate 95% confidence intervals. 116
4.7 Item invariance plots of the item difficulties of each of the four experimental conditions in Study 7: algorithm, computer, robot, and human. Each number represents an item. Red lines indicate 95% confidence intervals. 119

4.8 Visualization of participants' (a) person measures on the Predisposition to Anthropomorphize Scale, and (b) scores on the Waytz-instrument per experimental condition in Study 7. Whiskers represent 95% error bars. 121

List of Tables

2.1 Number of male and female participants per experimental condition in Study 2. 39

3.1 Examples of each of the artificial agents used in the emotion recognition task in Study 3. 63

3.2 The averaged response latencies, number of errors and anthropomorphism ratings for each of the expression-gaze combinations in Study 3. Standard errors between brackets. 65

4.1 Item difficulties (δ), infit- and outfit mean squares, human uniqueness and human nature of the Predisposition to Anthropomorphize Scale in Study 5. 96

4.2 Item difficulties (δ), infit- and outfit mean squares of the Predisposition to Anthropomorphize Scale in Study 6. . . 106

4.3 Item difficulties (δ), infit- and outfit mean squares of the Predisposition to Anthropomorphize Scale in Study 7. . . 115

4.4 Correlations between item difficulties in the experimental conditions in Study 7. 117

Appendices

The following appendices are included on the next pages:

A. Questionnaires
   This appendix contains all questionnaires used in this dissertation.

B. Washing tasks
   This appendix contains the washing tasks as used in Chapter 2.

C. Heating tasks
   This appendix contains the heating tasks as used in Chapter 3.

A

Questionnaires

On the following pages, these questionnaires are included in Dutch and English:

A1. Godspeed-instrument
A2. Waytz-instrument
A3. Game-evaluation
A4. Perceived intelligence
A5. Perceived agent knowledge
A6. Need to belong
A7. Agent-evaluation
A8. Predisposition to anthropomorphize

A.1 Godspeed-instrument

Below is the set of items of the anthropomorphism part of the Godspeed questionnaire (adapted from Bartneck et al., 2009), in Dutch and in English. Items had to be answered on a 5-point or a 7-point response format.

Dutch:

- "Nep - Natuurlijk"
- "Machine-achtig - Mensachtig"
- "Onbewust - Bewust"
- "Kunstmatig - Levensecht"
- "Beweegt houterig - Beweegt vloeiend"

English:

- "Fake - Natural"
- "Machinelike - Humanlike"
- "Unconscious - Conscious"
- "Artificial - Lifelike"
- "Moving rigidly - Moving elegantly"

A.2 Waytz-instrument

Below is the set of items of the anthropomorphism questionnaire adapted from Waytz, Morewedge, et al., 2010, in Dutch and in English. Items had to be answered on a 5-point or a 7-point response format.

Dutch:

- “In hoeverre heeft de avatar eigen gedachten?”
- “In hoeverre heeft de avatar intenties?”
- “In hoeverre heeft de avatar een vrije wil?
- “In hoeverre heeft de avatar een bewustzijn?”
- “In hoeverre heeft de avatar verlangens?”
- “In hoeverre heeft de avatar waarden en normen?”
- “In hoeverre ervaart de avatar emoties?”

English:

- “To what extent does the avatar have thoughts of its own?”
- “To what extent does the avatar have intentions?”
- “To what extent does the avatar have a free will?
- “To what extent does the avatar have a consciousness?”
- “To what extent does the avatar have desires?”
- “To what extent does the avatar have values and norms?”
- “To what extent does the avatar experience emotions?”

A.3 Game-evaluation

Below is the set of items of the game-evaluation questionnaire (adapted from Williams & Jarvis, 2006), in Dutch and in English. Items had to be answered on a 5-point or a 7-point response format.

Dutch:

- "In hoeverre werd u in het spel betrokken?"
- "In hoeverre voelde u zich geaccepteerd gedurende het spel?"
- "In hoeverre vond u het leuk om het spel te spelen?"
- "In hoeverre voelde u zich boos tijdens het spel?"
- "In hoeverre waren de andere spelers vriendelijk?"
- "In hoeverre waren de andere spelers gezellig?"
- "In hoeverre waren de andere spelers aangenaam?"
- "Ik voelde me geaccepteerd door de andere spelers."
- "Ik voelde me 'verbonden' met de andere spelers."
- "Ik had het gevoel dat er een connectie was met de andere spelers."
- "Ik voelde me als een buitenstaander tijdens het spel."
- "Ik voelde me tijdens het spel in staat om de bal te gooien zo vaak ik wilde."
- "Ik voelde me enigszins gefrustreerd tijdens het spel."
- "Ik voelde dat ik controle had tijdens het spel."

English:

- "To what extent were you involved in the game?"
- "To what extent did you feel accepted during the game?"
- "To what extent did you enjoy playing the game?"
- "To what extent did you feel angry during the game?"
- "To what extent were the other players friendly?"
- "To what extent were the other players sociable?"
- "To what extent were the other players pleasant?"
- "I felt accepted by the other players."
- "I felt a 'connection' with the other players."
- "I felt that I bonded with the other players."
- "I felt like an outsider during the game."
- "I felt capable of throwing the ball as often as I wanted during the game."
- "I felt somewhat frustrated during the game."
- "I felt that I was in control during the game."

A.4 Perceived intelligence

Below is the set of items of the perceived intelligence part of the Godspeed questionnaire (adapted from Bartneck et al., 2009), in Dutch and in English. Items had to be answered on a 7-point response format.

Dutch:

- “Onbekwaam - Bekwaam”
- “Onwetend - Heeft kennis”
- “Onverantwoordelijk - Verantwoordelijk”
- “Onintelligent - Intelligent”
- “Onverstandig - Verstandig”

English:

- “Incompetent - Competent”
- “Ignorant - Knowledgeable”
- “Irresponsible - Responsible”
- “Unintelligent - Intelligent”
- “Foolish - Sensible”

A.5 Perceived agent knowledge

Below is the set of items of the perceived agent knowledge questionnaire (adapted from Powers & Kiesler, 2006), in Dutch and in English. Items had to be answered on a 7-point response format.

Dutch:

- "In hoeverre heeft Femke bekwaamheid?"
- "In hoeverre heeft Femke kennis van zaken?"
- "In hoeverre is Femke intelligent?"
- "In hoeverre heeft Femke deskundigheid?"
- "In hoeverre is Femke betrouwbaar?"
- "In hoeverre is Femke nuttig?"
- "In hoeverre is Femke te vertrouwen?"
- "In hoeverre is Femke aangenaam?"

English:

- "To what extent does Femke have competence?"
- "To what extent does Femke have knowledge?"
- "To what extent does Femke have intelligence?"
- "To what extent does Femke have expertness?"
- "To what extent does Femke have reliability?"
- "To what extent does Femke have usefulness?"
- "To what extent does Femke have trustworthiness?"
- "To what extent does Femke have likability?"

A.6 Need to belong

Below is the set of items of the need to belong questionnaire (adapted from Leary et al., 2007), in Dutch and in English. Items had to be answered on a 7-point response format.

Dutch:

- “Als andere mensen me niet aanvaarden, dan heb ik daar geen last van.”
- “Ik doe heel erg mijn best om dingen te vermijden waardoor anderen mij zouden vermijden of uitsluiten.”
- “Ik ben zelden bezorgd of andere mensen om me geven of niet.”
- “Ik moet het gevoel hebben dat er andere personen zijn waar ik naartoe kan als ik problemen heb.”
- “Ik wil dat andere mensen me aanvaarden.”
- “Ik houd er niet van om alleen te zijn.”
- “Ik heb er geen last van als ik gescheiden ben van mijn vrienden voor een lange periode.”
- “Ik heb een sterke behoefte om ergens bij te horen.”
- “Ik heb het er moeilijk mee als ik niet betrokken ben bij de plannen van anderen.”
- “Ik voel me snel slecht wanneer ik weet dat anderen me niet aanvaarden.”

English:

- “If other people don’t seem to accept me, I don’t let it bother me.”
- “I try hard not to do things that will make other people avoid or reject me.”
- “I seldom worry about whether other people care about me.”
- “I need to feel that there are people I can turn to in times of need.”
- “I want other people to accept me.”
- “I do not like being alone.”
- “Being apart from my friends for long periods of time does not bother me.”
- “I have a strong need to belong.”
- “It bothers me a great deal when I am not included in other people’s plans.”
- “My feelings are easily hurt when I feel that others do not accept me.”

A.7 Agent-evaluation

Below is the set of items of the agent-evaluation questionnaire, in Dutch and in English. Items had to be answered on a 7-point response format.

Dutch:

- “Vond u de feedback in de wasmachine-taak nuttig?”
- “Vond u het aangenaam om feedback de avatar te krijgen?”
- “Vond u dat de avatar er knap uitzag?”
- “Vond u dat de avatar aantrekkelijk was?”
- “Vond u dat de avatar sympathiek was?”
- “Vond u dat de avatar vriendelijk was?”
- “Vond u dat de avatar nuttig was?”
- “Vond u dat de avatar betrouwbaar was?”
- “Vond u de interactie met de avatar aangenaam?”

English:

- “Did you find the feedback in the washing machine task useful?”
- “Did you find it pleasant to receive feedback from the avatar?”
- “Did you find the avatar handsome?”
- “Did you find the avatar attractive?”
- “Did you find the avatar sympathetic?”
- “Did you find the avatar friendly?”
- “Did you find the avatar useful?”
- “Did you find the avatar reliable?”
- “Did you find the interaction with the avatar pleasant?”

A.8 Predisposition to anthropomorphize

Below are all items of the Predisposition to Anthropomorphize Scale in Dutch. Items had to be answered with ‘yes’ or ‘no’.

Item (x’es indicate in which studies each item was used)	Study 5	Study 6	Study 7
De robot begrijpt een taal.	x	x	x
De robot begrijpt het morele dilemma.	x	x	x
De robot denkt na over het morele dilemma.	x		
De robot ervaart blijdschap.	x		
De robot gaat georganiseerd te werk.	x		
De robot heeft de intentie niemand te kwetsen.	x		
De robot heeft een vrije wil.	x		
De robot heeft gekozen om het morele dilemma op te lossen.	x		
De robot herkent emoties van anderen.	x	x	x
De robot is ambitieus.	x	x	x
De robot is doelbewust.	x	x	x
De robot is fantasierijk.	x		
De robot is in staat zich ongelukkig te voelen over het dilemma.	x	x	x
De robot is rationeel.	x		
De robot is zich bewust van de fysieke omgeving.	x	x	x
De robot is bewust van zichzelf.	x	x	x
De robot kan afstanden inschatten.	x	x	x
De robot kan op evenementen in de omgeving anticiperen.	x	x	x
De robot kan boos zijn.	x	x	x
De robot kan emoties van anderen begrijpen.	x	x	x
De robot kan kleuren waarnemen.	x		
De robot kan lopen.	x	x	x
De robot kan objecten oppakken.	x	x	x
De robot kan objecten waarnemen.	x	x	x
De robot kan obstakels vermijden.	x		
De robot kan pijn ervaren.	x		
De robot kan praten.	x	x	x
De robot kan raadsels oplossen.	x	x	x
De robot kan rekenen.	x	x	x
De robot kan springen.	x	x	x
De robot kan stemmen herkennen.	x	x	
De robot kan tevreden zijn.	x		
De robot kan zich in anderen inleven.	x	x	
De robot kan zien.	x	x	
De robot voelt zich verantwoordelijk.	x	x	
De robot voert met opzet acties uit.	x	x	
De robot ziet diepte.	x	x	

Below are all items of the Predisposition to Anthropomorphize Scale in English. Items had to be answered with 'yes' or 'no'.

Item (x'es indicate in which studies each item was used)	Study 5	Study 6	Study 7
The robot understands a language.	x	x	x
The robot understands the moral dilemma.	x	x	x
The robot things about the moral dilemma.	x		
The robot experiences happiness.	x		
The robot works in an organized manner.	x		
The robot has the intention not to hurt anyone.	x		
The robot has a free will.	x		
The robot has chosen to solve the moral dilemma.	x		
The robot recognizes others' emotions.	x	x	x
The robot is ambitious.	x	x	x
The robot is purposeful.	x	x	x
The robot is imaginative.	x		
The robot is able to feel unhappy about the dilemma.	x	x	x
The robot is rational.	x		
The robot is aware of the physical surroundings.	x	x	x
The robot is aware of itself.	x	x	x
The robot can estimate distances.	x	x	x
The robot can anticipate on events in the physical surroundings.	x	x	x
The robot can be angry.	x	x	x
The robot can understand others' emotions.	x	x	x
The robot can detect color.	x		
The robot can walk.	x	x	x
The robot can pick up objects.	x	x	x
The robot can detect objects.	x	x	x
The robot can avoid obstacles.	x		
The robot can experience pain.	x		
The robot can talk.	x	x	x
The robot can solve riddles.	x	x	x
The robot can calculate.	x	x	x
The robot can jump.	x	x	x
The robot can recognize voices.	x	x	
The robot can be satisfied.	x		
The robot can empathize.	x	x	
The robot can see.	x	x	
The robot feels responsible.	x	x	
The robot deliberately performs actions.	x	x	
The robot sees depth.	x	x	

B
Washing tasks

On the following pages, the set of trials as used in the laundry task in Chapter 2 are presented in Dutch and in English.

Dutch:

- "Was vier donkere wollen truien."[7]
- "Was een volle lading verschillend gekleurde vuile kleding, handdoeken en ander wasgoed dat in huis rondslingert."
- "Was een gemengde lading overhemden van polyester en katoen, katoenen sokken en verschillend gekleurd ondergoed."
- "Was de vloermatjes uit de badkamer."
- "Was een lading witte polyester vitrages."
- "Was een lading verschillend gekleurde katoenen handdoeken."
- "Was de kinderkleding die vuil is van het buiten spelen."
- "Was een lading vuile spijkerbroeken."
- "Was een lading geheel zwarte kledingstukken van verschillende stoffen, die een beetje vuil zijn."
- "Was een lading witte katoenen badhandoeken."
- "Was het kleed dat de bank bedekt en waar de hond op heeft gelegen."

The items below were only used in Study 2.

- "Was een lading wit beddegoed dat een week in gebruik is geweest."
- "Was een lading babykleertjes met etensvlekken en urinevlekken."
- "Was een lading kleding, zowel gekleurd als wit, waaronder katoenen t-shirts."
- "Was een lading witte katoenen handdoeken, die erg stinken."
- "Was een zijden overhemd dat u nodig heeft voor een feestje vanavond."
- "Was een kleine lading synthetisch damesondergoed."
- "Was een lading witte mannen-overhemden."
- "Was een lading vuile theedoeken met daarbij een vuil vloerkleedje uit de keuken."
- "Was een lading donkere synthetische kleding die redelijk vuil is."
- "Was een lading van verschillende kleuren en verschillende stof, welke slechts een beetje vuil zijn."
- "Was het witte tafellaken en de servetten van Kerstmis, met rode wijnvlekken."
- "Was een lading overalls."
- "Was een polyester dekbed."

[7] This item was used as practice trial in both studies.

English:

- “Wash four dark woolen sweaters.”
- “Wash a full load of variously colored dirty clothes, towel and other laundry that lies around the house.”
- “Wash a mixed load of polyester and cotton shirts, cotton socks, and variously colored underwear.”
- “Wash the floor mats from the bathroom.”
- “Wash a load of white polyester curtains.”
- “Wash a load of variously colored cotton towels.”
- “Wash the children’s wear that are dirty from playing outside.”
- “Wash a load of dirty jeans.”
- “Wash a load of black clothes of various fabrics that are a bit dirty.”
- “Wash a load of white cotton bath towels.”
- “Wash the rug that covers the couch on which the dog lied.”

The items below were only used in Study 2.

- “Wash a load of white bed linen that has been used for a week.”
- “Wash a load of baby clothes with food and urine stains.”
- “Wash a load of clothes, both colored and white, amongst which are cotton t-shirts.”
- “Wash a load of white cotton towels that smell very badly.”
- “Wash a silk shirt that you need for a party tonight.”
- “Wash a small oad of synthetic ladies underwear.”
- “Wash a load of white male shirts.”
- “Wash a load a dirty dishcloths together with a dirty floor rug from the kitchen.”
- “Wash a load of dark synthetic clothes that are quite dirty.”
- “Wash a load of variously colored and different fabrics that are just a bit dirty.”
- “Wash the white table cloth and the napkins with red wine stains that were used for Christmas.”
- “Wash a load of overalls.”
- “Wash a polyester duvet.”

C
Heating tasks

On the following pages, the set of trials as used in the thermostat task in Chapter 3 are presented in Dutch and in English.

Dutch:

- “Het is avond en je bent thuis. Het is buiten 3°C. Stel voor de verschillende vertrekken de centrale verwarming in.”
- “Het is middag en je bent niet thuis. Het is buiten 20°C. Stel voor de verschillende vertrekken de centrale verwarming in.”
- “Het is ochtend en je bent thuis. Het is buiten -5°C. Stel voor de verschillende vertrekken de centrale verwarming in.”
- “Het is ochtend en je bent thuis. Het is buiten 15°C. Stel voor de verschillende vertrekken de centrale verwarming in.”
- “Het is middag en je bent niet thuis. Het is buiten 8°C. Stel voor de verschillende vertrekken de centrale verwarming in.”
- “Het is zondagmiddag en je bent thuis. Het is buiten 19°C. Stel voor de verschillende vertrekken de centrale verwarming in.”
- “Het is middag en je bent niet thuis. Het is buiten 18°C. Stel voor de verschillende vertrekken de centrale verwarming in.”
- “Het is avond en je bent thuis. Het is buiten 6°C. Stel voor de verschillende vertrekken de centrale verwarming in.”
- “Het is avond en je bent thuis. Het is buiten 17°C. Stel voor de verschillende vertrekken de centrale verwarming in.”
- “Het is avond en je geeft vanavond thuis een feestje. Het is buiten 16°C. Stel voor de verschillende vertrekken de centrale verwarming in.”
- “Het is nacht en je ligt in bed. Het is buiten -1°C. Stel voor de verschillende vertrekken de centrale verwarming in.”
- “Het is nacht en je ligt in bed. Het is buiten 14°C. Stel voor de verschillende vertrekken de centrale verwarming in.”

English:

- “It is evening and you are at home. The outside temperature is 3°C. Set the temperature for the different rooms in the house.”
- “It is afternoon and you're not at home. The outside temperature is 20°C. Set the temperature for the different rooms in the house.”
- “It is morning and you're home. The outside temperature is -5°C. Set the temperature for the different rooms in the house.”
- “It is morning and you're home. The outside temperature is 15°C. Set the temperature for the different rooms in the house.”
- “It is afternoon and you're not at home. The outside temperature is 8°C. Set the temperature for the different rooms in the house.”
- “It is a Sunday afternoon and you're home. The outside temperature is 19°C. Set the temperature for the different rooms in the house.”
- “It is afternoon and you're not at home. The outside temperature is 18°C. Set the temperature for the different rooms in the house.”
- “It is evening and you are at home. The outside temperature is 6°C. Set the temperature for the different rooms in the house.”
- “It is evening and you are at home. The outside temperature is 17°C. Set the temperature for the different rooms in the house.”
- “It is evening and tonight you give a party at home. The outside temperature is 16°C. Set the temperature for the different rooms in the house.”
- “It is night and you're in bed. The outside temperature is -1°C. Set the temperature for the different rooms in the house.”
- “It is night and you're in bed. The outside temperature is 14°C. Set the temperature for the different rooms in the house.”

Acknowledgments

This dissertation would not have been here without the help and support of many people. Whenever I was in doubt about what I was doing or where I was going, there was always someone around to cheer me up and/or steer me in the right direction. I would like to take this opportunity and express my gratitude to all who helped making this dissertation possible.

First of all, to my supervisors Cees and Jaap. We did not always get along perfectly, but I would not have been where I am now without your help and guidance. Cees, I would like to thank you for giving me the freedom and possibility to develop myself on various levels. Your contributions were very valuable and I enjoyed the discussions we had about my work, science, and life. Jaap, whenever I was taking the job too seriously, you showed me that having fun is the basis of productivity.

To all my colleagues at HTI, for making me feel so much at home from the first day until the last. I'm very happy and grateful to have been part of this group. It has been a wonderful experience that I will always continue to cherish.

Daniel, you were not only my peer coach in the BKO program, but also a motivating and inspiring colleague for young PhD's like me. Your efforts in improving the scientific methods we use has made me look more critically at my own work as well.

Gerrit, thank you for giving me the opportunity to develop myself as a teacher. My involvement and responsibilities in the research methods course we taught together were an important part of my personal development, and encouraged me to learn more about various scientific methods as well.

Antal, our collaboration in this project, that ended up in Chapter 4 of this dissertation, was both pleasant and insightful. I have learned a lot about the pros (and cons, but mainly the pros) of using the Rasch model for measuring anthropomorphism (a word that you ironically cannot pronounce properly). The discussions we had about the topic have been extremely helpful to me.

Wijnand, both your interest in and helpful insights on my studies have enabled me to view them from different angles, helping me to see the bigger picture of the project. Yvonne, Raymond, Elena and Els, I really enjoyed our collaborations in supervising students and co-writing the (conference) papers. These collaborations made me broaden my scope as well, and have enabled me to publish on different topics than the ones presented in this dissertation.

The many private get-togethers with colleagues showed me that there is more than our scientific interests that bind us at HTI. We have been involved in afternoons of sports and games (‘meerkamp’), game nights, PhD nights, an angels game, and an uncountable number of Thursday- and Friday-afternoon drinks. Moreover, what started off as a funny smart-phone-gig at Wijnand’s inauguration party grew into an actual party band with an extensive portfolio (having performed at the Effenaar, the Trafalgar Pub, the Zwarte Doos and in multiple backyards).

Mieke, when we first worked together in 2009 on the day we ‘invented’ the game Flexcie, I had no idea that I would end up doing a PhD project, let alone that you would have such a valuable contribution to it. You were the first fellow PhD student who started using the Rasch model, and it was your description of the model that encouraged me to investigate the possibility for using it in my own research. I also want to thank you for being one of my seconds (‘paranimfen’) on the day of my defense.

Of all the colleagues at HTI, one of the smallest ones may have had the largest impact on my life, both work and private. Frank, you were much more than a colleague to me. Ever since the Spring School in Rotterdam, back in April 2011, we have been inseparable whenever (alcoholic) drinks were involved. Without you, my conference visits would not have included the legendary road trips to New York –ending with me using print screens on a laptop for navigation– and through California. As travel companion and roommate, you have given me the support I needed and were the best company I could have wished for. I am also very happy to have you as one of my seconds on the day of my defense.

To the people who I met abroad at one or multiple occasions. Special thanks to Friederike, Dieta, Birte, and Julia, with whom I have had great discussions about the meaning of the concept anthropomorphism. Most of my views on the topic stem from those discussions. Martina, our talks about the scientific methods in HRI, especially during our stay in Bielefeld, were entertaining in many ways and they helped me in critically viewing my own work.

To all the students who I (co)-supervised in one of the projects presented in this dissertation, especially Maartje van 't Sant, Diane Bouten, and Dana Rouschop. The projects we did together ended up as an important part of this dissertation. Our collaborations and your input have given me new insights in my own work.

Voor mijn vrienden en familie. Céryl, ik heb nog veel kunnen nagenieten van de tijd die we besteedden gedurende ons bestuursjaar bij Intermate, in Linköping, en in het PC-Koninkrijk, waar ik de grootste tijd van mijn project heb doorgebracht. Ook al kozen we elk een ander pad, onze discussies en samenwerking hebben aan de basis gelegen van mijn interesse in het doen van wetenschappelijk onderzoek, en hebben mij ertoe aangezet om te gaan promoveren.

Niels, Janno, Pablo, en Theo, de afgelopen vier jaar was ik misschien niet altijd even beschikbaar en bereikbaar, maar ik kon altijd op jullie rekenen om de stress rond het project te verlichten en te genieten van de vele andere mooie dingen in het leven.

Gabi, even though we met only a relatively short time ago, it feels as if you've been supporting me with the project for years. Thank you for being there for me in the most difficult moments, for providing occasional feedback, and for your patience with me during the final phase of the project. I hope to be of similar value for you during your project in the coming years.

Joyce, mam en pap, het was soms moeilijk voor me om uit te leggen wat ik nu precies aan het doen was, maar jullie hebben me altijd in al mijn keuzes ondersteund en zijn achter me blijven staan. Tot slot, mam en pap, hoe moeilijk jullie het zelf soms ook hadden, jullie zijn me altijd met veel liefde en warmte onvoorwaardelijk blijven ondersteunen.

Peter A. M. Ruijten
Eindhoven, April 2015

Curriculum Vitae

Peter A. M. Ruijten was born on 01-05-1984 in Stampersgat (Oud en Nieuw Gastel), The Netherlands. After obtaining his HAVO diploma in 2001 at Thomas More College in Oudenbosch, he studied Electrical Engineering at Avans Hogeschool in Breda. His HBO thesis concerned the development of a software tool for providing a real-time overview of machine failures inside a glass factory, and decreasing the number of palletizer machine failures. After obtaining his HBO degree, he studied the pre-master Human-Technology Interaction at the Eindhoven University of Technology. From June 2006 to May 2007 he was a board member of Study Association Intermate, after which he started the Human-Technology Interaction master program. From January 2009 to June 2009 he was an exchange student at Linköping University, Sweden, where he completed courses on Industrial Ecology and Sustainability. His master thesis concerned the use of social cues to influence littering behavior. He obtained his master degree in January 2011. From February 2011 he started a PhD project at Eindhoven University of Technology, as part of a long-term energy research project titled ‘Persuading households to save energy through smart agents’ in collaboration with Tilburg University and the Expert Centre on Smart Technology & Smart Living. During his PhD he was nominated for the best Late Breaking Reports paper at the 9$^{\text{th}}$ ACM/IEEE International Conference on Human-Robot Interaction in Bielefeld, Germany, and he was awarded with the Best Presentation Award at the 9$^{\text{th}}$ International Conference on Persuasive Technology in Padua, Italy.

Publications

Ruijten, P. A. M., Haans, A., Midden, C. J. H, & Ham, J. (2015). *Measuring perceived human-likeness: The development of the Predisposition to Anthropomorphize Scale.* Manuscript in preparation.

Ruijten, P. A. M., Midden, C. J. H, & Ham, J. (2014). *Ambiguous agents: On the influence of consistency of an artificial agent's social cues on emotion recognition, recall, and persuasiveness.* Manuscript submitted for publication.

Ruijten, P. A. M., Midden, C. J. H, & Ham, J. (2014). *Lonely and susceptible: The influence of social exclusion and gender on persuasion by an artificial agent.* Manuscript submitted for publication.

Ruijten, P. A. M., Ham, J., & Midden, C. J. H. (2014). Investigating the influence of social exclusion on persuasion by a virtual agent. In L. Gamberini, A. Spagnolli, & L. Chittaro (Eds.): *PERSUASIVE 2014, LNCS 8462* (pp. 191-200). Springer International Publishing Switzerland 2014.[8]

Ruijten, P. A. M, Bouten, D. H. L., Rouschop, D. C. J., Ham, J., & Midden, C. J. H. (2014). Introducing a rasch-type anthropomorphism scale. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction* (pp. 280-281). ACM New York 2014.[9]

Torta, E., van Dijk, E., Ruijten, P. A. M., & Cuijpers, R. H. (2013). The Ultimatum Game as Measurement Tool for Anthropomorphism in Human-Robot Interaction. In G. Herrmann et al. (Eds.): *ICSR 2013, LNAI 8239* (pp. 209âĂŞ217). Springer International Publishing Switzerland 2013.

Ruijten, P. A. M., Midden, C. J. H., and Ham, J. (2013). I didn't know that virtual agent was angry at me: Investigating effects of gaze direction on emotion recognition

[8]The presentation of this paper won the Best Presentation Award, based on the votes of the conference attendees.

[9]This paper was one of the 6 (out of 107) nominees for the best Late Breaking Reports award.

and evaluation. In: S. Berkovsky & J. Freyne (Eds.): *PERSUASIVE 2013, LNCS 7822* (pp. 192-197). Springer-Verslag Berlin Heidelberg 2013.

Ruijten, P. A. M., De Kort, Y. A. W., & Kosnar, P. (2012). Bridging the gap between the home and the lab: A qualitative study of acceptance of an avatar feedback system. In M. Bang & E. L. Ragnemalm (Eds.): *PERSUASIVE 2012, LNCS 7284* (pp. 251-255). Springer-Verslag Berlin Heidelberg 2012.

Ruijten, P. A. M., Ham, J., & Midden, C. J. H. (2012). The exclusion coefficient: The influence of social exclusion on sensitivity to the persuasive power of technology. Poster presented at SPSP 2012, San Diego, CA, USA.

Ruijten, P. A. M., Ham, J., & Midden, C. J. H. (2011). Social exclusion, anthropomorphism and persuasion: The influence of social exclusion on persuasion by an artificial social agent. Poster presented at ICSR 2011, Amsterdam, The Netherlands.

Ruijten, P. A. M., Midden, C. J. H., & Ham, J. (2011). Ambient persuasion needs goal-striving: The effect of goal activation on the persuasive power of subliminal feedback. Poster presented at Environmental Psychology 2011, Eindhoven, The Netherlands.

Ruijten, P. A. M., Midden, C. J. H., & Ham, J. (2011). Unconscious persuasion needs goal-striving: the effect of goal activation on the persuasive power of subliminal feedback. In *Proceedings of the 6th International Conference on Persuasive Technology: Persuasive Technology and Design: Enhancing Sustainability and Health* (p. 4). ACM, New York.

Colophon

This thesis was typeset using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. The cover of this dissertation is designed by Peter Ruijten & Harrie de Weert. This dissertation is printed and published by Proefschriftmaken.nl || Uitgeverij BOXPress, 's-Hertogenbosch.