universität wien

MASTERARBEIT / MASTER’S THESIS

Titel der Masterarbeit / Title of the Master‘s Thesis

"Empathy with Robots"

verfasst von / submitted by

Iris Nina Landsgesell, BSc

angestrebter akademischer Grad / in partial fulfillment of the requirements for the degree of

Master of Science (MSc)

Wien, 2018 / Vienna 2018

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 840
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Psychologie Psychology
Betreut von / Supervisor:	Ass.-Prof. Giorgia Silani, Privatdoz. PhD

Table of Contents

Introduction

Due to the growing practice of implementing artificial intelligence in technical agents and the consequently growing number of possibilities for people to interact with them in private and working environments, robots are increasingly participating in a variety of spheres of human life. This swift progression in the field of robotic technology has brought a new category of social interactivity to human society over the past decade. Specifically designed social and companion robots are created to establish satisfying relationships with their human partners and are already deployed in various fields of nursing and care taking.

As empathy is a fundamental component of social interaction, the potential and capability of robots to display emotional reactions towards humans is a much discussed and well established topic in a broad range of scientific disciplines, including the new field of robopsychology. Fundamental premises of reasonable human-robot-interaction (HRI) are conclusively based on robotic physical appearance and behavior, as recent evidence suggests that morphologic features in shape, motion and voice significantly contribute to the emotional perception and acceptance of robotic entities. Still, anthropomorphic thinking, the general tendency of projecting human mental states to artificial agents, seems to leave room for a high degree of abstraction in human-like features of non-human beings while still evoking emotional responses in a human observer.

This work investigates empathy towards robots and is divided into three parts. The first part of this work gives a general introduction to the subject area, such as the concepts of empathy and anthropomorphism and discusses fundamental processes of human perception related to human-robot-interaction. Within this theoretical overview, this work also addresses the actual discourse about artificial emotion and robot ethics and is presenting recent empiric findings of empathic reactions in social interactions of humans and robots. As there is only little systematic research to empathic reactions of humans towards robotic creatures, the major objective of the empirical part in this work is to investigate human empathy for robotic agents as a factor of human-like appearance in robotic morphology, both on an affective and a cognitive level. Furthermore, the study aims to investigate differences in empathic concern for robots that are related to personal traits and individual sociodemographics.

1 Theoretical Background

1.1 Anthropomorphism

“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, 1757)

The term "anthropomorphism" describes the human tendency to regard nonhuman entities as humanlike and thus project general human qualities onto them. In a psychological context, the term is generally used for the tendency to attribute human-like mental capacities such as feelings, wishes, desires, emotions and also reasoning to other entities or objects. (Epley, Waytz, & Cacioppo, 2007; Haslam, Bain, Douge, Lee, & Bastian, 2005). As anthropomorphism can contribute to the respectful handling of animals and nature, it is also deployed as a powerful tool for ecological concerns (Chan, 2013; Root-Bernstein, Douglas, Smith, Verissimo, 2013). Anthropomorphism also has an essential influence on human moral and ethical evaluations and supports pro-social behavior (Waytz, Cacioppo, & Epley, 2010).

With its long history in research and in the widely distributed contexts of religion, art and literature, the concept of anthropomorphism embraces a broad range of scientific approaches, which are also reflected by a myriad of explanatory models. For a better understanding, anthropomorphism has to be differentiated from other related terms, yet they can still contribute to the understanding of anthropomorphic thinking as a whole. Animism, for example, is a traditional concept whereby a spirit is allocated to the natural world, and is mainly found in religious contexts. As a part of the psychological developmental theory of Jean Paul Piaget (1929), animism signifies a tendency to project feelings, consciousness and intentions to inanimate objects as a magic way of thinking, which is evident in children from about two to seven years of age. Around seventy years after Piaget, Nass and Moon (2000) used the ancient Greek term "ethopoeia" in the context of new technologies, to describe the tendency of people interacting with media and computers in a social and socially adapted way. In contrast to anthropomorphism, where the attribution of humanlike features and states to inanimate objects can also occur on a conscious level, the authors are describing the phenomenon of ethopoeia as an automatic and unconscious mechanism.

Probably one of the earliest yet most famous examples of empirical research into the phenomenon of anthropomorphism is the experimental study of the psychologists Fritz Heider and Marianne Simmel (1944). The visual material deployed in this study is represented by a short animation film, showing three geometric forms, (two triangles and a circle), that move around and interact within the boundaries of a bigger square. According to the participants' narrations, after having watched the scenario, the graphic forms seemingly told a social story to them. Despite the extreme abstraction of the imaging, the presented figures were perceived as individuals interacting and acting in a motivated, intentional and social way. Emphasizing the importance of taking into account the complexity of the topic, Fisher (1991) created an elaborate theoretical framework with two main categories of anthropomorphism. One is termed "imaginative anthropomorphism", the other one "interpretive anthropomorphism". The latter describes the tendency to infer mental states of other species on the basis of their behavior and ascribing so called M-predicates (moral, personality, etc.) to them. Whilst stating that anthropomorphism is a historically and socially deeply ingrained phenomenon, the author is also a proponent of making a distinction between the act of attributing human characteristics to different entities and potential effects for interactions following that on a behavioral level. This distinction is also supported by Kiesler, Power, Fussell, and Torrey (2008), who found that people do apply general rules of human social interaction when interacting with robots, although they are aware of the fact that they are dealing with artificial entities.

Critical viewers of the concept of anthropomorphism on the other hand see it as an inadequate remedy for the interaction of humans and other species or non-living entities, mostly when there is no clear distinction to concepts of animism or religious beliefs. Some critics are referring to anthropomorphism as inappropriate in the context of the scientific community and research (Wynne, 2007). Nevertheless, the very early hypotheses of David Hume portraying anthropomorphism as a universal human tendency are subsequently reflected in current theoretical conceptions of different scientific approaches.

1.1.1 Social mechanisms of anthropomorphism

In an extended study on interdisciplinary orientation, Caporeal and Hayes (1997) were hypothesizing that anthropomorphism might present an automatic process, being represented by an evolutionary determined interspecies recognition system. As an alternative explanation to these automatic mechanisms Caporeal and Hayes (1997) are proposing that it also might be a sort of "cognitive default, which always occurs when there is no other reasonable explanation for the behavior of non-human entities". This assumption is also supported by de Waal (1999), conceptualizing anthropomorphism as a form of heuristic thinking and coining the term "heuristic anthropomorphism". Urquiza-Haas and Kotrschal (2015) suppose that both automatic (bottom-up) and reflective (top-down) processes are serving as a potential source of anthropomorphism. The former is representing effortless and fast responses that have domain-specific qualities, the latter is representing more cognitive mechanisms with domain-general abilities, like reasoning and inductive concluding. Inductive reasoning about the mental states of others can be based upon either categories or on similarities. (Sloutsky & Fisher, 2004, as cited in Urquiza-Haas & Kotrschal, 2015). These are also two main factors in conceptualizing Human-Robot-Interaction (HRI). On the one hand the physiological resemblance on the basis of movement, expression and form (similarity) and on the other hand the experience-related assignment to a category of objects, namely robots (category).

As a kind of reverse anthropomorphism, the social phenomenon of objectifying subjects and by that dehumanizing them might also contribute to a comprehensive approach of anthropomorphic thinking. Denying human qualities to other humans is termed infrahumanization and mainly appears between so called in-groups and out-groups. Haslam et al. (2005) are defining "humanness" with two distinct dimensions: “Human nature” includes attributes like depth, emotionality and warmth, and “human uniqueness” describes a culturally experienced ability for morale and includes concepts like rationality, civility and sensibility. As a possible way to identify psychological mechanisms underlying anthropomorphism that are referring to qualities beyond overt factors like physical appearance, language, etc. Zlotowski, Strasser, and Bartneck (2014) deployed the concept of dehumanization as a reverse phenomenon to anthropomorphism. Following Haslam et al. (2005), two factors of human-likeness (human nature and human uniqueness) were operationalized as dimensions of anthropomorphism and were realized in an interactive robot, displaying both cognitive and emotional traits. The assessment of the participants' evaluation after an interaction with the robot indicated that emotionality as a main effect had a clear influence on anthropomorphism, but intelligence did not contribute on a significant level. Hence, the authors derived that anthropomorphism is (at least) a two-dimensional construct, but with strong limitations from the potential of other factors contributing to the influence of emotionality and intelligence to the perceived human-likeness.

Also conjectured to be a basis for anthropomorphic thinking, is inductive inference based on perceived similarity and knowledge (Epley et al, 2007). According to this elaborate theoretical framework, with a high degree of perceived similarity either in behavior and/or morphology, people tend to employ egocentric knowledge and rely on quickly assessable information, when being confronted with varying species. If the target seems less similar to oneself, the process of inductive reasoning about the other is guided by stereotypes to a higher degree. Due to a growth of individual experience and knowledge about other beings, the possibility of using alternative schemata than one’s own identity increases and the individual tendency towards anthropomorphism declines as we grow older (Epley et al., 2007). Next to this factor of "agent knowledge", the authors are suggesting two other psychological factors of anthropomorphism: "effectance motivation" as the need to interact with others in a meaningful way and reducing uncertainty and "sociality motivation" as the need for social bonding. Motives of effectance as a driving force for anthropomorphism have also been examined by Waytz, Morewedge, Epley, Moneleone, Jia-Hong, and Cacioppo (2010) by creating a situation of unpredictability within a robots behavior, assuming it to be a strong trigger of ascribing mental states to objects. To increase the motivational aspect, the authors monetarily rewarded correct predictions about the robot’s future actions. Anthropomorphic thinking was assessed by a self-report scale, evaluating how much the subjects ascribed consciousness and intentions to the robot. Significant differences in anthropomorphism between incentivized and non-incentivized participants supported the hypothesis that efforts of understanding and obtaining control of a situation and regulating a social associate are significantly influential to anthropomorphic thinking.

1.1.2 Individual influences

Letheren, Kuhn, Lings, and Pope (2016) have empirically examined individual personal factors which are influencing anthropomorphic thinking, using the "Individual Differences in Anthropomorphism Questionnaire (IDAQ)" (Waytz et al., 2010) and psychometric measures to assess different personal traits. Results showed that openness to experiences, neuroticism and conscientiousness are personal traits that are significantly correlated with a higher tendency to anthropomorphic thinking. With reference to socio-demographic data, people of a younger age, being single and having a strong connection to animals have been identified to anthropomorphize to a higher degree (Letheren et al, 2016). Though consenting to the ubiquitous character of anthropomorphic thinking of human kind, Epley et al. (2007) are pointing out cultural, individual or situational factors within their complex matrix of potential determinants for anthropomorphic thinking. Due to current trends in care sector, considerations about situational and dispositional factors such as actual states of felt loneliness or social disconnection have to be particularly taken into account within the field of hospitalization and nursing, and might be used as an explanation for the growing and also successful use of robotic partners and therapy animals.

1.1.3 Psychobiological and neuroscientific aspects

"Sociality motivation" as a determinant of the theory of Epley et al. (2007) coincides also with findings from functional neuroimaging studies that have focused on identifying neurobiological correlates of anthropomorphic thinking. These findings agree extensively on brain areas, which are associated with higher mental processes of social cognition, similar to those displayed in interaction among human beings. Cullen, Kanai, Bahrami, and Rees (2014) examined the neural foundation of inter-individual tendencies in anthropomorphic thinking by using fMRI, a mentalization task including a robotic partner and the IDAQ. Results showed that an individual tendency for anthropomorphizing is being significantly reflected in the volume of the grey matter of the left TPJ, a brain area which generally accounts for processes linked to higher-order functions, e.g. reflecting about the mental state of others (Cullen et al., 2014). When asking subjects to participate in a social game that activates brain areas associated with the theory of mind, Hegel, Krach, Kircher, Wrede, and Sagerer (2008) identified increased neuronal activation in this area within higher degrees of human-like appearance of the technological interaction partner. Although the findings cannot be generalized due to the small sample size, they are undermining theories of anthropomorphic thinking as a process of higher cognitive order. Gazzola, Rizzolatti, Wicker, and Keysers (2007) have shown that people respond to robotic actions with the activation of similar brain areas that are also responding when observing biological movement as an automatic reaction. This finding also supports the theory of Caporeal & Hayes (1997) that quick responses in an automatic manner work not only for the same and other species, but also for other moving entities, when performing target-related motoric action.

Scheele, Schwering, Elison, Spunt, Maier, and Hurlemann (2015) identified a neurobiological basis of anthropomorphism by measuring pre-test endogenous levels of oxytocin that turned out to be a significant indicator for individual anthropomorphic tendencies. When watching graphic stimulus material based on the study of Heider and Simmel (1944), additional administration of intranasal oxytocin during the trial further increased the anthropomorphic attribution bias and significantly enhanced the individual tendency to attribute social meanings to a non-social stimulus. Thus, the existence of the body's own biological essentials does also promote the tendency for the perception of human-likeness in non-human agents.

1.1.4 Measuring Anthropomorphism

Despite a relatively broad consensus on anthropomorphism being a firmly established human phenomenon, standardized psychological measures for anthropomorphic thinking only exist in very small numbers, empirical studies and scientific reviews about anthropomorphizing robots are just as rare. An overview of the few well-implemented actual empiric measurements of human-robot interaction is provided by Bartneck, Kulic, and Croft (2009). The authors are pointing out that researchers in the field of HRI are generally confronted with a number of difficulties, such as the differentiation of various concepts like anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. Reflecting the complexity of the topics, several operationalizations for measuring these phenomena can be applied - including behavior, physiological reactions and personal attitudes, the latter with the greatest risk of receiving biased data. Powers and Kiesler (2006) evaluated varying factors of anthropomorphic appearance in a robotic head as an influence to peoples mental model of the robot and its given advices. Although the authors found direct effects of head shape and voice of the robot on the acceptance by the participants, the validity and generalizability of the findings are limited. What the authors are indicating as a limitation to the results was the elicitation of individual data of the participants as a potential influence on the findings, as the robot was very outgoing in verbal expression and by that might have merely reached out to persons with an extraverted personality.

Standardized questionnaires to measure individual tendencies of anthropomorphic thinking and perception of robots according to their physical features that are detached from a distinct robotic creature are very few and far between, probably also as a consequence of numerous difficulties in developing such validated scales. Chin, Ryan, Clark, Ballion, Dolezal, Shumaker, and Finkelstein (2005) have developed an instrument to measure anthropomorphic reactions in interactions between humans and non-human entities. The "Anthropomorphic Tendencies Scale" contains 208 items and includes one scale for acquiring data in the form of a self-report personal tendency for anthropomorphism and another scale for requesting the estimation of how other persons tend to anthropomorphize. By that, Chin et al. (2005) extracted four independent types of anthropomorphic tendencies: "(1) “extreme” anthropomorphic tendencies, (2) anthropomorphism toward a god or higher power, (3) anthropomorphism toward pets, and (4) inappropriate irritation with and anger directed at various non-human entities. According to the results, a distinction between the anthropomorphisation of living beings such as pets and non-living entities like technical machines seems highly probable, but further research addressing the influence of social desirability and actual behavioral tendencies when interacting with living and non-living entities is certainly indicated. Waytz et al. (2010) noted that the actual measurements to evaluate anthropomorphism are either too voluminous or do not address the concept of anthropomorphism sufficiently. For developing their "individual differences in anthropomorphism questionnaire" (IDAQ), the authors have taken the tendency to anthropomorphize as a stable individual trait next to factors of culture, experience or cognitive styles. As this measurement is part of the study conducted in this work, the questionnaire is described in paragraph 2.6.2.

In conclusion, numerous factors of situational, dispositional, developmental and cultural aspects contribute to the complexity of potential underlying mechanisms of the phenomenon of anthropomorphism, which challenges finding sufficient explanations and consent over the wide array of different scientific fields. As Caporeal und Heyes (1997) illustrate in their conclusive work "Why anthropomorphize?" it is impossible to ascertain their own diverging hypothesis, about if the underlying mechanisms of anthropomorphism are processes of "cognitive default", "overlapping species coordination system" or "a value-making activity of obligate social creatures". In the applied area of robotic sciences there is an essential need for further research and a requirement for approaches to empirically research human-robot-interaction as a function of anthropomorphic thinking.

1.2 Robots

Human and artificial creatures share a long history together, reaching back to the Greek and Hebrew culture, when myths about beings that are man-made emerged for the first time. A very early robot-like construction is said to be created by Archytas of Tarentum, a Pythagorean philosopher and mathematician, around 400 B.C.E. The wooden creature resembled a dove and the motivation for the construction was probably to better understand the essence of the birds' capability to fly (Huffman, 2012, p. 82). The first machine that was inspired by human anatomy was built by Leonardo da Vinci in 1495. It represented a kind of mechanical knight, made of steel and capable of basic movements like sitting and standing and independently moving its joints (Moran, 2007). From what we know today, the main reason for impelling the construction of robotic machines was - much like today - to serve humans and help them with their strenuous work. The core significance of those mechanical creatures is reflected in the verbal origin of the word “robot”. It originates from the Slavic word “rabota”, which means “servitude” and was created by the Czech author Karel Čapek (1920/2004). In his novel “Rossum´s Universal Robots” from 1929, the author already conceptualized robots as more than solely technical machines, and described them as beings hardly to distinguish from humans.

With further progression in developing robotic machines, another application field of the mechanic creatures was found. With the intention to entertain people, several “toy automata" were built and became operational in public and private locations. The idea of robotic creatures, which are not only working devices but social partners to people, is by this time an essential element of actual Science-Fiction. From that moment on, when robots were not to be seen as mere functional instruments anymore, a fundamental change of their roles was taking place. This change increasingly positioned robots in social contexts and as social partners of human society. Robots are not only found in the field of labor as co-workers of humans, but also in explicitly social contexts as companions and partners in medical, therapeutic and individual fields of application. Hence it seems that the human kind gained a new partner for social and communicative interaction, as the implementation of artificial intelligence in social robots is incrementally becoming a part of our everyday social life. Fundamental issues of emotional aspects in social relationships between robots and humans have brought theorists from different areas to the scene and gave rise to a new psychological field, namely Robopsychology. In Austria, a research unit of this new scientific field of psychological research has been established in 1996 within the Ars Electronica Futurelab.

An important issue in the in the novel scientific field of robotics is the question of an appropriate classification according to bodily features of technical agents. Robots with a very strong humanlike appearance are generally termed as "androids", whereas those who reflect human physiognomy in a strong, but still technical and mechanical way are mostly referred to as "humanoid" (Schlobinski & Siebold, 2008, p.85). De Graaf (2016) even proposes to establish a new ontological category for social robots, as they can neither be classified as animate, nor as complete inanimate objects. A rather practical approach to describe robotic entities from the perspective of a human interaction partner would be a categorization according to different mental models of the users in robotic application. Breazeal (2003) created different categories of robotic entities, clustering them as "tools", as "cyborg extensions", as "avatars" and as "sociable partners". Each of these categories represents varying grades of autonomy that are also reflected by the variation of physical overlaps with the human body. Limb prostheses, for example, would be categorized as a cyborg extension and despite blurring the boarders of robot and human, have become a widely accepted and integrated amplification of the human body by now (Breazeal, 2003).

1.2.1 The Uncanny Valley

Robotic physiognomy according to their features and morphology does occur in a very broad scope of different shapes. It is by now largely understood, that enhanced human-likeness of any technical devices is facilitating acceptance by their users and is evoking increased social behavior. Amid this general phenomenon, there is a distinct manifestation of robotic phenotypes, where this effect is not only suspended, but reversed into its opposite. The Japanese roboticist Masahiro Mori was one of the first to explore the aspect of humanlike aesthetics in the field of robotics. As an often cited thesis it is often used to explain irritations in HRI, which are suspected to occur due to high anthropomorphic appearance of robotic creatures. Mori's (1970) theory states, that the grade of acceptance of humans towards robots depends on two distinct features of the robots: 1) the degree of human-likeness and 2) the motion of the robot. Generally a robotic machine is more favorable for its users, the higher the degree of similarity with humans is. However, this curve of growing acceptance is drastically falling off on a distinct point of very high similarity to humans, just to go back to positive evaluation, when a maximum of human-likeness is reached. Within this delineated zone the "uncanny valley" is situated. According to Mori (1970), robotic movement is further enhancing the emotional effect towards the machines, putting moving androids straight into the uncanny valley.

Despite of being widely spread in the scientific community and having a strong impact on empirical and theoretical work dealing with robotic morphology, the theory of the Uncanny Valley is also regarded as hypothetic and scarcely approved scientifically. During the last decades, the theory of the uncanny valley has undergone some critical reviews and was partly confirmed and partly disproved. Fang-Wu (2016) at least confirmed one of the theses of Mori, as the subjects in the study clearly preferred robots with a moderate but still distinguishable degree of human-likeness to those with strong android looks. Regarding motion though, the results contradicted Mori's theory, as movement increased the acceptance of the robots among the participants instead of weakening it. When presenting video material of the robots instead of pictures, movement did not have a significant influence of the ratings in likeability and flattened the curve of the uncanny valley, suggesting that movement might be associated with higher levels of social and physical acceptance (Fang-Wu, 2016). Based on the morphologic aspect of Mori's theses, Ferrari, Paladino, and Jetten (2016) found that strong anthropomorphic appearance evoked most disapproval as it was experienced as "threat to distinctiveness" by the participants. Robots with high human-likeness were further experienced to be most threatening for humans. The theory of the uncanny valley was confirmed insofar, as the impediment in distinction between androids and humans generated a clear discomfort among the participants (Ferrari et al., 2016).

Researchers around Hiroshi Ishiguro (Bartneck, Kulic, & Kroft, 2009) were able to position their study on the fine line between a real human being and a perfect replication of it, by utilizing android Geminoid HI-1, which is an exact replication of his developer Ishiguro. Three levels of human-likeness were realized by deploying the developer Ishiguro himself, his android robot and the slightly modified android, to reduce his very human-like appearance. The participants had a short interaction including a conversation with all of the three entities; the interactions were conducted either with movement or without. The collected measures in according to "likeability" did not confirm Mori's theory. Although the participants were able to differentiate between the android and the human, they reported no differences in the likeability of both and movement of the entities did also not influence the factor of likeability. Allocating social partners to varying categories is a common part of human interaction, as a possible explanation for the results the authors are presuming that people might apply different standards for human and non-human agents and by that the ratings of "likeability" might not be comparable between the different categories. According to the authors, another limitation to the findings might be language-related. The Japanese word Mori used was throughout translated as "familiarity", but according to Bartneck et al. (2007) the original idea might have been lost in translation, as maybe "likeability" is a more appropriate term. As the authors found a positive correlation between their applied categories of "familiarity" and "likeability", they however plead for regarding anthropomorphism as a multi-dimensional construct. MacDorman (2006) also assumed that human-like appearance might not be the only factor to explain the acceptance or rejection of robots and implemented scales of "eeriness" and "familiarness" next to "human-likeness" when letting people rate different robotic entities. Results showed that with the same level of rated human-likeness the perception of robots as being familiar or eery did clearly deviate. This finding is also reflected by the every-day phenomenon of perceiving dolls or stuffed animals as familiar because of anthropomorphism, but eery at the same time.

With emphasizing the lack of empirism in Mori's concept of the uncanny valley and bringing up the general obstacles of empiric research in this field, Misselhorn (2009) chose a more philosophic approach to the topic. According to the author, androids as inanimate entities might trigger the innate human aversion to anything reminding humans of their own mortality. As this does not happen in any and every way, Misselhorn (2009) additionally supposes aesthetic features as influential for the eery effect of androids, as they trigger a classification in human categories, which yet at the end cannot be related to. The uncanny valley might therefore represent a kind of bridge between life and death (Misselhorn, 2009). Pursuing this approach, this chapter concludes with a consideration of Sigmund Freud who within his theory of "negative" aesthetics described the ambivalence between familiarness and unfamiliarness with the associated uncertainty as a source of uncanny feelings (Freud, 1919).

1.2.2 Human Perception of Robots

With increasing embodiment of artificial intelligence, questions of robotic design regarding their anthropomorphic appearance are gaining rapidly in importance. Next to the robot's outward appearance, behavioral aspects and possibilities for interaction, are influencing factors of the acceptance of social robots and do shape the interaction of robotic and human partners in a significant way. Following, I would like to introduce two prominent theories of psychological perception processes that are also related to social cognition. Due to only limited empiric data about how people are apprehending robotic or android entities on an affective level, the following concepts might contribute to the exploration about how technical entities are perceived by humans by means of their physical appearance.

1.2.2.1 Gibson, affordance, design

In the context of robotic behavior and interaction with them, the main question is, if humans perceive actions of non-human beings in a way similar to those of their own species, respectively other human beings. Technical machines as functional units usually have a well defined purpose and are mostly being perceived by their potential users according to their function.

The perception of objects according to their supplied possibilities was first described by the psychologist James Gibson. In the "Affordance Theory", Gibson (1977) described the connection between objects and their specific clues representing action possibilities on one hand, and on the other hand the action capabilities of their potential users, which are related to each other and the shared environment. With their theoretical framework for the use of autonomous robots, Sahin, Cakmak, Dogar, Uguer, and Göktürk (2007) have extended Gibson's model by implementing three varying perspectives of potential perception, namely those of the agent, the environment and the observer. This extension is diversifying some aspects of HRI. With a recent shift in robotic technology by putting robotic agents from a technical object into the position of an acting subject, affordance theory is mainly applied to enable robots as agents to ideally interact with the affordances of their environment and utilizing affordances of their ecology in flexible and autonomic ways. In HRI, with a human in the perspective of the agent, robots however do either portray a functional machine or a potential social partner, mainly depending on the outward appearance and affordances related to it. Although theoretical considerations related to affordance theory are frequently applied as useful concepts in applied science of HRI and robotics, still the multi-facetted term of affordances is actually "both inspirational and hazy", as Sahin et al. (2007) put it.

1.2.2.2 Action and perception

Models of action and perception describe a system of common coding and shared representations in the human brain that are activated, when both an action is performed by oneself and when actions of other entities are observed. For the perception of humans and other mammal species, this mechanism is often described, but the possible degree of abstraction while still perceiving entities or objects in similar ways to humans, is barely investigated. Some researchers are reporting that the human mirror neuron system is only activated when being confronted with distinct human clues representing biological motion, but not with mechanical or technical entities (Tai, Scherfler, Brooks, Sawamato, & Castiello, 2004). Meltzoff and Brooks (2001) do describe the process of shared acts to become shared minds on the part of developmental psychology, with taking in aspects of both innate equipment and individual experiences to make sense of human acts in the children's environment. The emphasis of the authors on the premise of "being like me" for understanding bodies and minds of others, would support the idea of a growing understanding of robotic entities, as the reproduction of biologic movement and expression is continuously progressing in robotic technologies. Although the reported degree of intensity in activation is varying in actual literature, recent studies in HRI do correspond in their findings of robots being capable of evoking positive responses in the human motor system, just as humans do. (Oberman, McCleery, Ramachandran, & Pineda, 2007; Oztop, Franklin, Chaminade, & Cheng, 2005; Press, Bird, Flach, & Heyes, 2005).

Models of social cognition associated with the mirror neuron system, are generally assuming, that the essential mechanisms for understanding others actions through observing them, is the reception of the motor act as target-orientated and not as random or meaningless (Bouqet, Shipley, Capa, & Marshall, 2011). Bisio, Sciutti, Nori, Metta, Fadiga, Sandini and Pozzo (2014) have examined the influence of varying motion patterns of human and robotic hands and the differences in processing motion that is either object or non-object directed. Effects of motor contagion became apparent in any of both conditions (with target and without), but the effect however was not only mediated by "objected-directness", but also by "kinematic" aspects of robotic movement. Human subjects only showed effects of resonance when the moving stimulus ranged within the motor repertoire of the human observers, with a significant influence of the velocity of the movement (Bisio et al., 2014). Therefore, not only morphology and mobility, but also the resemblance of biological structures according to aspects like pace, seem to be mandatory conditions for activating motor contagion within humans. Investigating the effect of specific aspects of robotic motility in peoples' perception, Kupferberg, Huber, Helfer, Lenz, Knoll, and Glasauer (2012) found a higher effect on the distinct configuration of a robotic joint than of an overall humanlike appearance. This supports the significance of the execution of quasi-biological movements for triggering processes of human motor resonance. Wykowska, Chellali, Al-Amin, and Müller (2014) found that reaction time during a motor task significantly depended on a prior visual clue, but only in the means of a content-related congruency of the clue. The morphology of the prime (human or robotic hands) itself had no significant effect on reaction time. For a better generalization of their findings, Wykowska et al. (2014) were proposing to further realise various degrees of similarity of robot morphology to human morphology. However, creating systematic variations of the specific features of robotic entities along an anthropomorphic spectrum is still one of the biggest challenges for Robopsychology and scientific research in this field.

Despite clear evidence for the ability of artificial creatures and motion to have a strong and commensurable impact on human perception, biological beings still seem to elicit stronger representations in the human brain than artificial systems. Still, effects of learning through augmented contact with technical agents have to be considered. According to the hypothesis of associative learning, that describes the development in matching observed actions and executions by learning, Press, Bird, Flach, and Heyes (2005) found a constant increase in the compatibility effect of reacting to robotic hands during an ongoing course of studying the perception of robotic movements. However, it can be assumed, that intensified contact and ongoing exposure to robotic beings, processes of learning and habitualness will have an impact on human perception processes. It is thus rather safe to assume that robotic motion patterns that at the very moment seem unnatural and deviating from human movements, will progressively become comprehensible for the human perceptual system.

1.2.3 Social aspects of Human-Robot-Interaction (HRI)

In addition to these systematically measurable human responses to robotic action, on a social level, people do show a clear tendency to relate to artificial beings (Nass & Moon, 2000). Discussing the necessity of specific theories for HRI, Krämer, von der Pütten, and Eimler (2012) note, that as long as an interaction between humans and robots seems to be sufficiently social, it follows similar patterns as human-human-relationships. Thus, it is considerable, that human motives of social bonding like the "need to belong" and "social exchange and equity" do also count for relationships between humans and robots in comparable manners (Krämer et al., 2012). Although the employment of social robots in medical or supporting care is highly discussed, their positive impact on emotional and behavioral states of patients by bonding with robotic therapy animals is well documented by now. The robotic seal named "Paro" is often quoted as a vivid example for the deployment of a robotic animal to decrease feelings of loneliness during geriatric nursing (Robinson, MacDonald, Kerse, & Broadbent, 2013). With a humanlike appearance, the social robot named "Telenoid" is operated by a visually unseen assistant that communicates through the speakers in the body of the robot. The robot was developed by Hiroshi Ishiguro, a prominent Japan roboticist, and was tested on patients with early stages of Alzheimer’s disease. As reported, both of the persons interacted actively with the robot, communicated with it and stated that they experienced the interaction with "Telenoid" as positive and natural. Although younger persons, who got into contact with the robot, perceived it as quite scary, the phenomenon of the uncanny valley did not show with the elderly test persons, neither at the beginning nor during the interaction with "Telenoid" (Yamazaki, Nishio, Ishiguro, H., Norskov, Ishiguro, N., & Balisteri, 2012). Despite the benefits of therapeutic robots, mostly argumented by limited space and interdictions of real animals, there are also reverse effects coming along with the application. When people lack social relationships and human contact, companion robots might even lead to higher feelings of loneliness and disappointment than having no social encounter at all, due to their still noticeable deficits in standing in for "real" social experiences.

In conclusion, it can be stated, that the general motivation for getting in contact with artificial creatures and the way the contact is being established, does under certain circumstances not seem to differ fundamentally from how people interact socially with human conspecifics. Krämer et al. (2012) were pursuing the question, whether specific social rules are applied in contact with unanimated beings and if there is requirement for a special theoretical framework for human-robot-interaction. Despite existing differences in HHI and HRI, the authors are convinced, that "now and in future there will be more similarities between human-human and human-machine interactions than differences".

1.2.3.1 Do Robots have Emotions?

The human ability and the tendency to attribute feelings and intentions when interacting with other humans, is mainly facilitated by the general assumption that all conspecifics do possess a fundamentally similar psychological architecture. Artificial emotion does pose a major challenge for human routines of social interaction. Not only do incremental realizations of humanlike embodiment blur the clear distinction between humans and robots, in the field of social robotics it has become quite common to implement emotional components in the computational mental architecture of a robot.

Nass, Steuer, and Tauber (1994) showed that when people socially interact with computers, they are attributing human social schemata to them, despite the awareness that inanimate beings do not possess any personality or consciousness that can be related to in emotional or social ways. Technical progress in artificial emotional systems is by definition prompting questions about emotional relations between humans and non-humans and about the "nature" of emotion generally. For a basic consideration of computational emotion and due to the enormous complexity of the topic, the following thoughts about potential robotic emotion are limited to basic concepts of psychological emotion theory.

With a long history in psychological emotion theory, the James-Lange Theory of Emotion states, that physical experiences within the body are regarded as not only an indispensable requirement for emotion, but can be equated with them (James, 1884). Actual theories that are putting bodily experiences into the center of emotional experiences are mostly represented by researchers in the area of the somatic feedback theory (see e.g. Prinz, 2004). According to theoretical foundations of embodied emotions, emotions are direct consequences of physical sensations, which do not require any higher cognitive processing. Applying these considerations on computational models of emotion, it might be said, that due to embodiment and corresponding sensory constitution, robotic creatures would then be in the position of at least experiencing basic affective emotions. The combination of sensorimotor technology, which enables robots to receive and process external physical stimuli and transferring them into changes of bodily states, and the cognitive structure to identify these "somatic" changes, do fulfill the qualification of experiencing emotions on an elementary level. Emotion as a regulatory process, both on social and on individual levels is also an important consideration in current emotion theories, particularly in psychobiological theories of emotion. On the basis of a hierarchical organization of emotional processing, basic emotions as primary processes are described by Panksepp and Watt (2011) to be innate and hard-wired with distinct neuronal circuits and have a strong influence on behavioral and motivational factors by contributing to processes of self-regulation and self-protection. This form of emotional regulation is a well established practice in social robotics at this time, reflecting human regulatory processes through basal emotional experiences. Equipping robots with an emotional system, which is signaling danger, does subsequently lead to adapted, self-protective behavior. Ziemke and Lowe (2009) are also emphasizing that homeostatic functions of emotional processing apply equally to human systems and to artificial ones. The framework for embodied synthetic emotion calls on core principles of traditional emotion theories and defines emotion as "(a) closely connected to embodied cognition, (b) grounded in homeostatic bodily regulation, and (c) a powerful organizational principle - affective modulation of behavioral and cognitive mechanisms - which is useful in both biological brains and robotic cognitive architectures " (Ziemke & Lowe, 2009). At the core of their comprehensive model of an artificial cognitive and emotional system, is the principle of homeostasis, realized through a complex and intertwined combination of affective and cognitive components. The model also refers to internal simulations of emotions, as described by Damasio (1999) with the term "as-if body loops".

As a representative of a somatic feedback theory combined with findings of modern neuropsychology, Damasio (1999) describes emotions as physical states which derive from bodily reactions to the outer world and stay on an unconscious level. The detection of these bodily states ("emotions") by distinct brain structures is then labeled as "feelings", which subsequently may become overt to the mind by "feeling a feeling". With the distinction between "emotions", "feelings" and the concrete experience of the latter, namely "core consciousness", Damasio (1999) is emphasizing on the significance of high-order cognitions such as self-awareness and consciousness for experiencing the entire spectrum of the emotional system. Following these considerations, it seems conceivable to equip robots with "emotions" and "feelings", but with no capability of further experiences procession in the means of Damasio's "feelings of feelings". Although present research in social robotics is already working on self-reflective technical agents with introspective abilities, at the time it doesn’t seem reasonable to assume at this time that systems of AI are capable of experiencing "qualia", as this distinct subjective quality of emotional experience is often termed. The empirical evaluation of quality and structure of subjective emotional experiences does still present a major challenge also in humans' psychological research, depicting a problem inherent to the scientific investigation of concepts related to consciousness or self-identity. From the perspective of computational affective modeling the concept of "feeling" is anyways a "problematic and ill-defined construct" as Eva Hudlicka (2008) puts it. Hudlicka (2008) has set up a framework for modeling emotions in a computational architecture with focusing on emotion generation and emotion effects, both main processes within the field of application. By drawing upon four fundamental constitutive aspects of emotion (behavioral/expressive, somatic/physiological, cognitive/interpretive and experiential/subjective), the author is also emphasizing on the multi-modal direction of emotion and the importance for effectiveness in believability of cognitive-affective architectures. Hudlicka (2008) further emphasizes on the importance of implementing cognitive-affective architectures in synthetic agents to enhance their effectiveness and believability for the users.

Research in robotic emotion and implementing it to robotic technologies has initially started with a focus on emotional expression. In the late 1970s, Ekman and Friesen (1978) identified six basic emotions that are supposed to be distinguishable from each other through their (facial) expression, building the basis for a "Facial Action Coding System". The premises of the theory of basic emotions that emotions help to cope with fundamental challenges of the environment and are an essential part of fundamental social interaction, can also be applied to shape robotic emotional expression. With humans, prototype emotional states or basic emotions, namely "seeking, fear, rage, lust, care, panic/grief and play can be evoked by an artificial activation of subcortical networks of the brain" (Panksepp & Watt, 2011). Furthermore, actual findings in neuroimaging support the assumption that a distinct and distinguishable neuronal pattern can be identified for each basic emotion (Vytal & Hamann, 2010). Regarding the functional structure of a human brain as an analogy to a computational system, basically any neural structure can be reproduced on a computational level. Regarding robotic emotions as merely technical simulations of emotion, there is clear substance to the idea of implementing basic emotions in robotic beings - at least on an expressional level. Already twenty years ago, Cynthia Breazeal (1998), then researching in the "MIT Artificial Intelligence Lab" in Massachusetts, has implemented emotional expression in a robot named "Kismet". This early and by now very prominent example of a sociable robot, was designed with the ability of expressing six basic emotions and was equipped with visual, sensual and proprioceptive sensory inputs to perform social interactions with humans in an emotional context. In course of the rapid development in computational emotion, by now formal concepts of artificial empathy are evolving, including complex fundamental parts of human empathy, such as the distinction between self and other and cognitive functions like perspective-taking (Asada, 2015; Damiano, Dumouchel, & Lehmann, 2015).

Still, what Breazeal and Brooks (2005) have pointed out more than ten years ago remains relevant up to this day. The robotic researchers in the MIT media lab stated that, they see no realistic option of implementing an appropriate emotional structure in artificial intelligence, as long as there is no deep insight into the nature of human emotion. Still, possible ethic aspects of generating computational models of emotion in artificial beings are an important topic in robot psychology. Nitsch and Popp (2014) address the question, if it is justifiable at all, to equip creatures with an emotional structure, as this ability per se also includes aversive and negative experiences. A relevant question therefore is not only if robots *do have* emotions, but if robots, after all, *need* emotions.

1.2.3.2 Do Robots need emotions?

Emotional expression of robots for their future relationship with humans and social integration is, without question, of high importance. Interestingly, emotional attributions to robots from a human perspective occur even on minimal emotional clues of robotic expression. Accordingly, Adolphs (2005) is fundamentally questioning the necessity of equipping robots with a personal emotional structure, when it might be sufficient to make them act with humans in a way that make believe, that they actually *have* emotions. Social robots as described by Breazeal & Brooks (2005) thus are mainly constituted in a way that allows them to categorize emotional situations in social interactions with humans, followed by emotional expression and also an appropriate behavioral reaction.

A rather functional, pragmatic position in the question of a demand for robotic emotion is also represented by Rosalind Picard, the director of the Affective Computing Research Group at the MIT. For Picard (2003) the need of equipping robots with an emotional structure lies mainly in the purport of organizing HRI as little as possible frustrating for the user. Next to the ability of recognizing human emotions, this aim can be realized most likely by abilities like attention shifting, decision-making processes and intelligent and flexible acting and reacting. Although all of these capabilities are strongly linked to the human understanding of emotion, Picard (2003) would suppose not to use the term "emotion" in a robotic context, for these abilities are merely functional standards for a more pleasant usage of robotic agents.

Apart from the important and ongoing discourse about the necessity and functionality of robotic emotions, research and findings in the very new scientific field of "artificial emotions" do not only improve the interaction of social robots with their human partners but might also contribute to the research of human emotion, as representants of artificial psychology are delineating (see e.g. Wang, Xiu, & Lu, 2016)

1.3 Empathy

"Empathy is often defined as understanding another person's experience by imagining oneself in that other person's situation: One understands the other person's experience as if it were being experienced by the self, but without the self actually experiencing it. A distinction is maintained between self and other." (Hodges & Myers, 2007, p.296)

1.3.1 Simulation Theory and Theory Theory

Over the last decades, a large number of terminologies, empirical approaches and theoretic definitions for the concept of empathy have been issued in the scientific fields of psychology, philosophy, neuroscience and others, recently also supported by new technologies of neuroimaging. Although there is no definition or explanation of complete consistency, there seems to be broad consensus of a bi-directional orientation with cognitive and affective components. Regarding theoretical concepts of empathy, a range from concepts of emotional contagion to highly cognitive activities such as mentalizing can be found. At the core of the general ability of reading others mind with a long scientific tradition there are "Simulation Theory" and "Theory Theory". Both theories describe the ability to gain insights into the inner mental states (desires, beliefs, etc.) of others, but are widely being regarded as opposing concepts. Simulation Theory says that one represents the mental activities of others by mental simulation, for example by generating similar activities and processes in oneself (Cruz & Gordon, 2002). According to Shanton and Goldman (2010), experiences in form of representations of mental states are written into a persons' individual mental structure from the early childhood on and are constantly used, when interacting with other people or when observing social situations. Simulation Theory is therefore describing the capability of inferring mapping premises and associated inferences of other people to own premises and inferences, supporting the ability to understand social courses of action. Although Theory Theory is also assuming that the premises for understanding others do onset in early childhood, it is adopting a more abstract approach of how people make sense of social situations and manage to appraise the minds of others. Addressing the developmental aspect, it is often assumed, that children do develop theoretical concepts about the mental states and intentions of others, which are further developed and refined through experience and observation of the social environment. This procedure is also linked to the concept of the "Theory of Mind" and is often described in analogy to a scientific process, similar to empiric-based theories that are utilized towards other people and their minds (Gopnik & Meltzoff, 1997; Gopnik & Wellman, 1994). With an ongoing debate in the scientific community for decades, however, there is still no consensus according to these fundamental concepts about why and how people understand the mental states of others.

1.3.2 Mirror neurons and neuroscience

With the discovery of the mirror neurons by Giacomo Rizzolatti in the late 1980s, Simulation Theory has experienced an upturn by findings on the base of neuroscience. Being activated not only when performing an action but also when observing one and by that providing a simulation of other people's actions, mirror neurons are considered as neuronal base for the ability of action understanding (Rizzolatti & Craighero, 2004). It is furthermore assumed that this kind of "mirroring" or simulation, as described in action-perception-models, is not only an explanation for the process of understanding actions and intentions. Also in affective contexts, similar neuronal structures are activated - both, when observing or imaging an emotional relevant situation of another person and when actually experiencing an emotional state by oneself. Gallese, Keysers, and Rizzolatti (2004) describe this phenomenon in the context of social situations as internal representations that are creating an "as if" emotional state in the brain of the observer. Thus, the function of mirror neurons is probably making a substantial contribution to the affective component of empathy, as their "…activity seems to be nature's way of getting the observer into the same 'mental shoes' as the target – exactly what the conjectured simulation heuristic aims to do” (Gallese & Goldman, 1998). These representations and their associated somatic and automatic reactions are assumed to play an important role as potential precursors to empathy, Preston and De Waal (2002) are referring to it as the proximate basis of empathy. Relating to mechanisms of action and perception, Preston (2007) has developed the "perception-action model for empathy" (PAM). Pursuant to classic theories of motor contagion, this model is conceptualizing shared representations as a core function to the ability of experiencing similar feelings to those of others. These representations are, likewise to the theory of motor inference, automatic and connected to somatic responses but can only be congruent to a certain degree, depending on factors inherent to object and target. With a rather comprehensive approach, the model understands empathy as a process, which contains proximal and ultimate components and affective and cognitive routes and includes also non-human or even abstract entities, like e.g. the environment. In contrast to other theories, where the distinction of experiences of self and other presents a fundamental part of empathy, Preston (2007) does not insist on a distinction between projection and empathic reactions. As shared representations within empathic processes are inevitably anchored in the subject, for Preston (2007), the extent of overlap and therefore the experienced empathy is a variable of similarity, familiarity and past experience.

Based on neuroscientific research, Singer (2006) clearly distinguishes between the ability of understanding others cognitive states or understanding and sharing others emotional states and labels these abilities "empathizing" and "mentalizing". Regarding the functional mechanisms, mentalizing includes only non-affective mental states, empathizing on the other hand refers to the idea of understanding emotion by sharing affective states of others through own bodily experiences and is therefore associated with the concept of mirror neurons. Still, according to Singer (2006), empathizing contains a wide range from those rather automatic processes to higher cognitive abilities, like e.g. perspective taking. Although in functional brain imaging the neuronal circuits of both empathizing and mentalizing are reflected as distinct and determinable neuronal circuits, they are also intertwined and interacting (Singer, 2006). This inclusive view clearly contributes to multi-level construct of empathy, frequently demanded in recent literature.

Albeit the popularity of the concept of mirror neurons and their widely spread publicity in the scientific community, the accentuation on their influence on empathy also raises critical voices. With a comprehensive analysis of the actual scientific findings, Lamm and Majdandžić (2014) are prompting to not overestimate the significance of mirror neurons as a substantial or even an exclusive source of empathy. Next to somatosensoric and motor mechanisms with related automatic processes, Lamm and Majdandžić (2014) are pointing out the crucial role, that higher cognitive processes can also play as a path to "affective" empathic reactions. Critically questioning the role of shared neural activations, the authors are underlining the relevance of mechanisms such as the theory of mind or mentalizing as a source of empathy, especially in the context of abstract imagination in absence of a real-life situation. For another critical view of mirror neurons as an exclusive explanatory model of social cognition or empathy also see Jakob & Jeannerod (2005).

Reconsidering the division of empathy into cognitive and affective components, Lamm and Majdandžić (2014) propose to use the term "cognitive perspective taking" for understanding cognitive mental states in difference to understanding or sharing affective states of others. As the latter would be linked to the concept of empathy as understanding and sharing affects, it becomes quite apparent, that there is only limited consensus about empathy-related denotations and terminology in present literature.

1.3.3 Mentalizing, mind-reading, empathizing

Terms to describe the phenomenon of understanding and sharing the affects of others, like perspective taking, mentalizing or mind-reading, are often used with no broad consent in current literature.

Referring to the Simulation Theory, Shanton and Goldman (2010) use the term "mind reading" for the ability of assigning mental states to a target and propose two distinguished mechanisms according to their depth in processing. "Low level mind reading" is described as an automatic and implicit process that is often not even accessible for consciousness and refers primarily to very simple movement patterns or basic emotions. In contrast, "high level mind reading" does refer less to one's own representations and more to the application of more elaborated information and imagination (Shanton & Goldman, 2010). This concept does reflect the dichotomy of the consensus division of empathy into cognitive and affective aspects, on which there seems to be a broad consensus within the scientific community and present literature by now. Coevally, a distinct segregation is anyhow controversially discussed and there is growing consensus about the importance of an integrative view of affective and cognitive components of empathy, as already emphasized by Davis (1983) decades ago The actual approach of the scientific community to conceptualize empathy as a complex and multilevel-construct is probably best reflected in comprehensive empathy models that embrace findings from psychology, philosophy and neuroscience. Taking into account this complexity, Decety and Jackson (2004) have developed a functional model of empathy, which describes the complex combination of various distinct and parallel processes with levels of lower and higher cognitive order. According to this model, inhibitory processes are serving as attenuation factors for the self-perspective and thereby facilitating to also take the perspective of the other person involved. This function of "mental flexibility″ is paving the way for higher cognitive processes, as effortful and controlled components of empathy. In the context of simulation processes, represented by automatic and non controllable mechanisms, two other important components of this model are "emotion regulation" and "self-awareness". These abilities serve as a prominent function in avoiding an undesirable merge of self and other, which is crucial for the avoidance of aversive experiences in the observer and do play a fundamental role in the following sub-domain of empathy.

1.3.4 Empathy for pain

Empathic responses resulting specifically in the context of painful situations of other people are generally termed as "empathy for pain". Recent neuroimaging studies have shown that experiencing pain (self-pain) and observing pain (other-pain) leads to activation in similar brain networks. These common neural circuits and shared representations while observing a painful situation of another person, without having received any dolorous stimulus oneself, have widely been considered as neuronal correlates of empathy for pain (Decety & Sommerville, 2003; Jackson, Meltzoff, & Decety, 2005). Combining neurobiological findings with more abstract ideas of psychological processes in a homogenous framework does however present a challenge transdisciplinary researchers are constantly facing. In the context of having examined the implication of shared activations in empathy, Lamm, Bukowski, and

Silani (2016) are referring to those fundamental concepts pointedly as "language of the brain" and "language of the mind". The veracity of equating the activation of overlapping brain areas in self-pain and other-pain with shared representations was examined by Rütgen, Seidel, Silani, Riecanský, Hummer, Windischberger, et al. (2015). By combining two methods of self-report-data and neurofunctional imaging, the authors compensated for the problem that is inherent in the system when using only one empirical method to apprehend complex psychological processes. Investigating aspects of self and other in empathy for pain, first hand pain experience was realized through electrical stimulation, pain of others by letting the participants observe other participants being exposed to the same painful stimulus. During the trials, placebo analgesia gel was used to reduce self-pain experience and by that exploring the influence of reduced first hand pain experience on the participants' empathy for pain. While considering to be under the influence of analgesia, the self-reported scores of empathy decreased and the fMRI measures showed a significant reduction of activation in the brain areas connected to empathy for pain. Interestingly the decrease in self-reported empathy did show on both affective and cognitive levels. In a second experiment the authors used an opioid antagonist to annihilate the effect of the primal analgesia placebo condition, which set the self-experienced pain and also the pain empathy back to initial valuation. Taken together, these findings suggest that neural activations and representations of first-hand experiences of pain and pain empathy in both cognitive and affective components for the pain of others do have a clear correspondence (Rütgen et al, 2015).

Still, it has to be noted, that even though the activation of brain areas in self-pain and other pain coincide to a significant degree, the activated patterns are not fully congruent. Painful experiences in an individual do include perceptual-sensory and emotional-affective components, summarized under the term "pain matrix". As several neuroimaging studies report, observing pain of others does however not touch upon the complete spectrum of the pain matrix. Activation in the somatosensory and the sensorimotor cortex is specific to the very own experience of pain and does not show when observing or actively taking the perspective of a pain-affected other person, which is activating brain areas that are related to higher cognitive processes, such as the right temporo-parietal junction. The increased degree of somatosensory activation might serve as a crucial distinction between self and other, since self-perceived pain can by no means be the pain of another person (Singer, Seymour, O'Doherty, Kaube, Doland, & Frith, 2004). According to Decety and Lamm (2006) empathy seems to be a dynamic interplay of forces between automatic processes that are the foundation for the ability to share emotions with another person (bottom-up) and the ability of controlling those affective reactions by cognitive regulation mechanisms (top-down). When being confronted with a painful situation of another person, and the activation of negative shared emotional representations is not regulated by putting focus on the self-perspective, personal distress is likely to occur (Lamm, Batson, & Decety, 2007; Decety & Lamm, 2009). Self-other-distinction thus provides an important contribution for the individual to preserve empathic feelings and not letting the experience turn into a painful of its own (Jackson, Brunet, Meltzoff, Decety, 2006). Furthermore, potential incapacities in overcoming emotional egocentricity may involve danger of shifting empathic reactions towards feelings of distress, as Lamm, Bukowski, and Silani (2016) are emphasizing.

1.3.5 Social aspects of empathy

In the present literature, there is widespread consensus that the emotional impact of rejection, social exclusion, or the loss of a close relationship are comparable to experiences of physical pain. On a neurobiological level, recent findings indicate an activation of congruent neural mechanisms, both for physical and for social pain (Eisenberger, Lieberman, & Williams, 2003; Macdonald & Leary, 2005). Analog to experiencing physical pain, in social emotions the distinction of self-perspective and other-perspective also decide on empathic concern or personal distress (Ruby & Decety, 2004). To include the complete spectrum of potential sources of empathy, in the present study, the interactions with the robotic partners were realized, both on a physical and on a social level.

Subsumed under the concept of perceived similarity, factors of common identities, attitudes, behaviors and social categories are often assumed to be another contributing aspect to empathy and prosocial behavior (Krebs, 1975). Social psychology does explain the influence of social categorization and the phenomenon of segmentation between an in-group and an out-group as a moderating factor for empathy and also for associated motivational factors (Dovidio, Johnson, Gaertner, Pearson, Saguy, & Ashburn-Nardo, 2010). Annoting the lack of empiric evidence for the similarity-hypothesis of empathy, Batson, Lishner, Cook, and

Sawyer (2005) conducted an experimental study to identify the role of similarity, which however did not display a direct influence of similarity on empathy. In contrast to the widely fielded hypothesis of a general effect of similarity on empathy, the authors were therefore assuming a more indirect or moderating relationship between similarity and empathy. A significant correlation between an empathic tendency with objects quite dissimilar to the participants (e.g. puppies) lead Batson et al. (2005) to the assumption that "nurturing" as a general human tendency to protect, might possibly contribute to empathic reactions. This hypothesis also punctuates considerations about an ultimate base for human empathy (Preston & de Waal, 2002) and contributes to current issues in social robotics about the morphologic design of companion robots and how it influences acceptance and the interaction with their users. Especially in social or learning contexts, childlike robotic appearance, which does not necessarily resemble human nature, is widely favored and does generally show a higher grade of acceptance among their human partners.

1.4 Empathy with robots

Interactive robotic toys as vivid examples of "fictional empathy" (Fuchs, 2014) or "direct empathy" (Tisseron, Tordo, & Baddoura, 2015) have a tradition of more than thirty years in human society. Furthermore, people giving names to household appliances or cars, are clearly indicating that the phenomenon of integrating non-human entities into everyday social life exists. But not only in a close, private context do people show signs of emotional relationships with robotic machines. This phenomenon is widely spread over different social contexts, even in rough social environments like the military service that growingly deploys robotic agents. Reports and research of incidences of soldiers getting emotionally attached to their robotic war comrade are increasing (see e.g. Carpenter, 2016) and have also gained great importance in military strategies. The story of an U.S. military robot, which was developed to detect land mines, is a prominent example of this phenomenon. When the robotic machine got seriously damaged during a trial, the colonel in duty ordered to stop because he could not stand the tragedy of the scenery and rated the whole process as inhumane (Taylor, 2012, p.xii). Notably, also robots with a very high degree of anthropomorphic abstraction can become targets of emotional concerns, as the reactions to the crash of a space probe of the "European Space Agency" recently showed. When the orbiter "Rosetta", which has neither limbs, nor any face-like features, crashed and got stuck on a comet in 2016, media was full of emotionally charged reports and prevalently there was even talk of the death of the space probe (BBC, 2016).

Even though these incidents from everyday life speak in favor of a clear human tendency to anthropomorphize, empiric evidence of empathic reactions toward robots or androids are still scarce. With a prominent study and a rather early empiric experiment in the field of human-AI-interaction Nass, Steuer, and Tauber (1994) showed that people have a tendency to treat computers politely and do apply social rules on the interaction with them. Although the users stated, that they are well aware of the fact that computers do not possess any kind of personality whatsoever, their interaction implied several distinct features of socially adapted interaction amongst humans. Furthermore, when the computer underlined its autonomous identity by denoting itself as "I", higher degrees of likeability were reported. Reeves and Nass (1996) labeled this phenomenon as "media equation", confirming that people tend to treat computers as social actors.

A theoretical framework for potential targets of empathy including entities from real to fictional or virtual entities was modeled by Fuchs (2014), labeling the different forms of empathy as "intercorporeal empathy", "extended empathy" and "fictional empathy". "Intercorporeal empathy" corresponds with general concepts of empathy, linked to the influence of embodiment and shared representations, or as the author puts it, an "interaffectivity", including a bodily-affective communication. "Extended empathy" resembles cognitive concepts of empathy, with focus on the imaginative and the as-if-character of this cognitive operation. "Fictional empathy" is found to the least extent in the current literature. By blurring the border between reality and virtuality, this idea is presenting a kind of connection between the other two forms of empathy and relates to non-living creatures, whose either physical features and movements or their comprehensible/intentional behavior are capable of evoking empathic reactions (Fuchs, 2014). This latter conception of empathy might well be applied to robotic creatures, as they represent border crossers between inanimate, technical and biological entities. In contrast to a simple technical machine, a robot is capable of moving independently and due to the cognitive architecture of robotic agents, they are also capable of acting in an intentional way, which is often regarded as fundamental part of understanding others.

In their theoretical model of empathy with robots Tisseron et al. (2015) are taking in the aspect of identifying with artificial agents on a physical level, according to the phenomenon generally known as "body ownership". In contrast to "auto-empathy", which is exclusively referring to the own self, being represented by an avatar, a robot as an embodied avatar might become a projecting surface for subjective states and feelings of the user, guided by the reactions and behavior of the robot, similar to a relationship when parenting. Being related to, but still distinct from the self, a robotic avatar can thus become a target of "direct empathy", as described by Tisseron et al. (2015). In the field of application in e.g. the care sector, this kind of projection can especially support the acceptance of robots as caregivers, as robotic agents might trigger feelings of belonging to own experiences and does not display endangerment. As the third and fourth form of empathy with robots, Tisseron et al. (2015) define "reciprocal empathy" and "intersubjective empathy". These forms of empathy are of particular interest to the present work, as emotional expressions of robots can lead to ascribing own emotions to them and even evoke the assumption that robots might be capable of taking the perspective of an interaction partner (Tisseron et al., 2015).

1.4.1 The role of embodiment

Cynthia Breazeal, a pioneer in sociable robotics, already pointed out more than ten years ago, that due to their embodiment, robots are capable of putting a physical impact to their human partners and do also influence interactions with them (Breazeal, 2002). As new generations of computer games are increasingly using embodied avatars, this offers a new research opportunity in the field of HRI. When performing a user study with "SenToy", an interactive control device, Höök (2008) observed, that watching the avatar on the screen resembling the emotional input given by the players through "SenToy", led to a reciprocal emotional influence of the human and the robotic agent. Höök (2008) labeled this effect as "affective loop", which is representing the base of an embodied affective system. According to the author, affective experiences of a fictional other can turn into personal and "real" experiences of the human interactor by means of identification. Nishio, Koichi, Hidenobu, and Ishiguro

(2013) were monitoring an effect of human emotion regulation through having a conversation with a teleoperated android that was capable of displaying different facial expressions. With observations similar to the "affective loop", Nishio et al. (2013) described the process as "body ownership transfer". The found effect occurred to a higher degree when the android was actively controlled by the participants, than when the participants were communicating passively, without any personal interaction with the android. The psychological phenomenon of "body ownership" or "body transfer illusion" is frequently described in the literature by the "rubber hand illusion" and is successfully applied within clinical psychology and psychotherapeutic treatment. In the area of applied science, the phenomenon of taking the perspective of a virtual character or incorporating an embodied avatar can be of great help for patients to improve and correct the perception of their own bodies, especially with persons being affected by conditions of body image disturbances.

Embodiment and the specific characteristic of body features do significantly contribute to the attribution of mental states to inanimate agents. Krach, Hegel, Wrede, Sagerer, Binkofski, and Kircher (2008) let people believe to play social games with AI with different manifestations of embodied human-likeness. Although the participants were actually interacting with solely human partners, the perception of their social partners according to their alleged feelings, intentions, etc. was significantly varying according to the degree of human-like appearance. Vaes, Meconi, Sessa, and Olechowski (2016) have pursued the gripping question of how far the degree of abstraction within the bodily features of a robotic body can go while still generating anthropomorphic thinking and evoking emotional reactions. By using vegetables, which have no capability of movement and are nowhere humanlike shaped per-se, any kind of goal-directed behavior and any form of mirroring effects were foreclosed a-priori. A relationship between the participants and the vegetables was only established by giving the latter a human name or a trait-describing adjective. Observing the vegetables in painful and non-painful situations, EEG results showed that naming the vegetables was a sufficient clue of humanness to evoke empathic reactions within the participants. Results however were only significant with participants who were prior identified as "high humanizers" (measured with an implicit association test) and persons with high scores in self-perceived empathy. Furthermore, within the last time course of the EEG measure, an activation of brain areas was found, that are linked to higher cognitive processes and also to cognitive forms of empathy. These results suggest that empathic reactions toward non-human entities, even at a very high degree of abstraction, are possible and may moreover also include a non-automatic component. Suzuki, Galli, Ikeda, Itakura, and Kitazaki (2015) let people observe painful situations of human as well as robotic hands. The EEG measures of the participants showed an activation of the P3 component for both human and robotic hands, but while the observation of human hands already caused a significant neural reaction in the chronologically early phase (ascending phase), an activation watching the robot hand only took place in the later (descending) phase of the P3. Since the P3 wave is linked to top-down processing of empathic reactions according to Suzuki et al. (2015), the observed effect might display an obstacle in human perspective-taking of non-human entities to a full extent.

Kate Darling, a research specialist at the MIT Media Lab, states that there is a high tendency for human beings to project human features to non-human creatures and thereby identifying with them, as people nowadays are intensely primed by science fiction and everyday culture regarding robots as potential social partners. Darling moreover is convinced that the human brain is biologically hardwired "to project intent onto any movement in our physical space that seems autonomous to us" (Darling, 2017). Following this statement, it can be assumed that through increasing affinity with different kinds of robotic beings in the future, social proximity will in all probability further alter the relationships of humans and robots and will lead to presumably stronger effects in humans relating to robots.

1.4.2 Intentional mistreatment of robots

A video clip of a doglike robot that was vigorously kicked by a human and still trying to make his way by moving along became viral on social media only recently. Despite the fact, that the "mistreatment" was carried out by one of its developers to test the motion capabilities of the highly evolved robotic dog, people all over the world got emotionally upset about this scene. After numerous complaints to PETA, the animals' rights organization finally had to issue a statement about the incident. Although the group noted, that there is an uneasy feeling about the pictures, they also clarified that no actual animal abuse was taking place (CNN, 2015). Similar resentful reactions also occurred with a video in which a humanlike robot was kicked and abused by its developer as a part of a functionality test. Anthropomorphic appearance but also the perceived viciousness and intentionality of the enacted mistreatment seem to be crucial factors for an empathic reaction of the viewers.

Creating stimulus material to empirically evaluate emotional reactions to robots with systematically controlled levels of humanlike features is a general problem of psychological research in the field of robotics, resulting in only few empiric studies about robot abuse. Confounding variables of morphology, movement, speech or behavior may always have an uncontrollable impact on the results. Despite these difficulties, Riek, Rabinowitch, Chakrabarti, and Robinson (2009) conducted a study to evaluate the direct influence of human-likeness on the potential of evoking empathic feelings in humans. By using four different robots with increasing anthropomorphic appearances and exposing them to neutral or abusive treatment, the authors found a significant correlation of the participants' empathic reactions with the particular grade of human-likeness. Addressing methodical aspects, it should however be noted, that the empathic reaction was only captured by a single question, which might not meet the requirements of empathy as a multidimensional concept to the full effect. Ward, Olsen, and Wegner (2013) examined the influence of intentionally harming different entities (a vegetative patient, a robot and a dead person) which are generally perceived as having no consciousness. Results indicated that a distinct intention to harm an inanimate object leads to perceiving it as a victimized being, which significantly increases mind attribution to the subject. Hoenen, Lübke, and Pause (2016) showed that negative behavior and intentionally aggressive treatment towards non-living entities elicit strong automatic reactions, represented by an activation of the human mirror neuron system. Mediated by an explicit malicious social interaction, even a functional machine like a vacuum cleaning robot was perceived by the noninvolved and observing participants as a victimized social entity followed by empathic concern.

The probably most noteworthy study about mistreating a robot was the replication of the famous Milgram Experiment, conducted by Bartneck and Hu (2008), with the only difference of a robot being tormented instead of a human partner. Although all participants administered the maximum amount of electricity to the robot, it was evident, that it triggered stress and feelings of discomfort within the subjects. When the robot expressed verbal and physical signs of suffering during the trials, the participants displayed obvious feelings of compassion with the robot. As limitation for the results however, Bartneck and Hu (2008) indicated a potential ceiling effect by not providing the possibility of stronger electric shocks, which would probably have lead to an even higher degree of abusiveness. As an additional limitation it has to be pointed out, that all of the participants were students or employees of a technical university. This regular exposition and affiliation to technical agents might have altered the findings by means of either a stronger or weaker effect on the conducted abusive behavior. In a second experiment, adding the influence of anthropomorphic thinking, Bartneck and Hu (2008) tested the potential influence of the robot's intelligence by instructing the participants to destroy the robot with a hammer after having socially interacted with it. Results showed that the aversion to kill the machine was significantly correlated to the perceived intelligence of the robot. Limitations for the experiment were addressed by the authors as the obvious low cost of the robot, for this might evoke less qualms to destroy it, compared to a more elaborated creature. Less on a physical level of destruction, but more on a psychological challenge to "kill" a robot, Bartneck, van der Hoek, Mubin and Mahmud (2007) examined the influence of intelligence and added "agreeableness" as a social factor to the perceived animacy of a robotic machine. After having played a cooperative game with it, the participants were prompted to switch the robot off, which it disapproved strongly and begged for further enduring. Despite the fact that all participants decided to shut the robot down, there were differences in the time span to do so, revealing a significant influence of the intelligence of the robot and the reported sympathy for it.

A rare systematic evaluation of emotional reactions to the mistreatment of robots has been implemented by Rosenthal-von der Pütten, Kramer, Hoffmann, Sobieraj, and Eimler (2014) by using a standardized measure for positive and negative effect (PANAS), psycho-physiologic responses and self-reported data. After a personal interaction with a robotic dinosaur, the participants were witnessing the mistreatment of the zoomorphic robot. Interaction with the robot induced significantly higher scores in both physiological arousal (electrodermal activity) and in self-reported emotional measures, than only watching the robot passively. However, the authors acknowledged that, as a dinosaur is not part of the actual human environment, it might display an unclear stimulus according to the concept of human similarity or familiarity. Furthermore, it shall be noted that factors of cuteness and schemata of childlike characteristics which are mostly inherent in pet toys, might also be an influence to the found effects and might significantly influence the found effects in the emotional reactions. However, the application of animals as prototypes for behavioral and physiological components in robotic design is frequently implemented as it seems to enhance the likeability of robotic partners.

1.4.3 Robot Ethics

As the empiric studies presented in this work and numerous other occurrences indicate, mistreatment of robotic creatures emotionally concern people. Accordingly, social intercourse with non-animate creatures is inevitably invoking issues of ethical and moral issues and is therefore calling for a critical consideration.

Human values of ethical concern are mainly linked to the projection of human characteristics in terms of mental states, feelings and experiences to a reference subject, be it another human being, a plant, an animal or an inanimate object. In a large survey, Gray, Gray, and Wegener (2007) asked over 2000 participants to assign eighteen different factors of mental capacities to different entities, e.g. a human, an animal, a dead person, God and a sociable robot. The authors statistically then identified two distinct factors of mind perception, which the authors termed as "agency" (self-control, emotion recognition, thought) and as "experience" (hunger, fear, desire, personality, pride, etc.). While babies and animals were evaluated as beings with high "experience" and low "agency", the participants rated other adults and themselves as beings with high scores in both, agency and experience. Remarkable for this present work, is the finding that according to the respondent's estimations, a robot has no experience at all, but an average degree of agency, significantly more than a baby or a chimpanzee and about the same as a young girl.

It seems that people show clear tendencies of ascribing mental states and emotional experiences to robotic entities and as a consequence, ethic aspects concerning creatures outside the human and animal world are of increasing importance. In 2006, the European Robotics Research Network (EURON) has released the "Roboethics Roadmap Book" (Euron, 2006). This profound document includes nuanced definitions of robots as a new species with bodies and minds, and embraces various ethic dimensions in a multitude of potential fields of robotic usage. The application moral standards on an entity is above all a question of ascribing mental processes, such as self-knowledge, affective states and autonomy to it. In this respect, the EURON has formed different categories, based on opinions of society, technicians and researchers: "Robots are nothing but machines", "Robots have ethical dimensions", "Robots as moral agents" and "Robots, evolution of a new species". These categories reflect the range of potential ethic aspects of robots, with artificial entities as both technical machines with a need of ethic restrictions in their application, but also as active agents and the eventuality of robots to become moral instances themselves. This aspect is of special importance, as the capacity of self-awareness in artificial intelligence is, by the current state of scientific knowledge, no mere fiction anymore. Under the term of "Artificial Consciousness", different scientific fields from psychology to technology and philosophy actually put strong efforts in defining the central components of so called "conscious machines" and therefore deriving a consequential legal state to them.

Questions of robot ethics have already also reached political domains. In 2015, the European Parliament sent a comprehensive recommendation to the commission of civil law regulation in the field of robotics. The draft resolution rests upon already existing legal frameworks referring to questions of liability or safety of robots in working contexts, and is highly extended by addressing those aspects in civil fields. Furthermore, topics reaching from the future of the labor market with robots potentially taking a great number of jobs, to aspects of data privacy, human dignity and the necessity of a worldwide legal regulation of robotics, are addressed (Europäisches Parlament, 2015). Of particular interest for the present work, is the specific contentual orientation in trying to define social relations of humans and robots. As in some cultures emotional bonding with robotic entities is by now a socially accepted part of human relationships, the fundamental ideological orientation in the European society is still laying emphasizing on a clear distinction between robots and humans.

2 Methods

2.1 Research Question and Hypotheses

The key research question of this study is to examine differences in cognitive and affective empathy towards robots as a function of physical features according to human-like appearance. Furthermore, the study seeks to examine the influence of personal traits and sociodemographic data on individual measures of empathy for robots. The hypotheses that were tested are as follows:

H1.1: Cognitive empathy towards robots with more humanlike physical features will be different than towards less humanlike robots.

H1.2: Affective empathy towards robots with more humanlike physical features will be different than towards less humanlike robots

H2: Empathic reactions towards robots are correlated to individual personal traits.

H3: Empathic reactions towards robots are correlated to individual sociodemographic data.

2.2 Participants

Potential participants to the study were contacted via social media, in persona and via email and invited to participate to a study that is exploring human perception of robots. The participants were informed that their data would be collected anonymous and would remain confidential. There were no specific criteria for inclusion or exclusion, besides a minimum age of 18 years and understanding German language sufficiently.

Altogether 64 participants participated in the study, 12 of them had to be excluded because of the rate of missing values, resulting in a sample size of (N = 52). Thirty-three (63.5%) participants declared themselves to be female, 19 (36.5%) to be male. The age of the participants ranged between 25 and 69 years, (M = 41.63, SD 11.189) with 53.8 % being 40 years or older. Having lived together with a pet in their childhood was reported by 41 (78.8%), 14 (26.9%) participants do live together with a pet at the present time. To have at least one technical device that was given a name, was reported by 11 (7.1%) participants, 40 (78.4%) have not named a technical device, one (0.6%) did not answer the question. 33 (63.5%) of the participants have a university degree as highest educational achievement.

With varying professional backgrounds and different age cohorts, it can be assumed that the participants are socialized with the use of computers or robotic machines and being familiar with the concept of artificial intelligence to also varying degrees.

2.3 Stimulus Material

The participants watched 32 videos clips with a medium length of 3.64 seconds (SD=0.8), which have been produced by PhD Giorgia Silani and Iris Landsgesell. The video material used for this study was shot with a SONY PMW-100 (XDCAM HD) camera and edited with the software Adobe Premiere. The material was filmed in the rooms provided by Otto Bock Healthcare Products GmbH, with Mr. Markus Schachinger, MSc., as the companies' technical expert on site. The technically highly evolved and electrically controllable hand prostheses were presented to the participants as a body part of a robotic creature. With three prosthetic hands, varying manifestations of human-likeness were realized, following a list of the four levels:

Level (1): A minimalistic metallic hook with two grapplers

Level (2): A hand with five fingers and humanlike joints, but pronounced technical and mechanical features

Level (3): The same prosthetic hand as in level (2), but covered with a flesh-colored silicone sleeve

Level (4): A real human hand

Images of the three robotic hands, representing the levels of human-likeness (1, 2 and 3) and the human hand (4) are depicted in Figure 1, 2, 3 and 4. Throughout this paper, the term "hook" will refer to level (1), the term "robot" to level (2), the term "hand" to level (3) and the term "human hand" to level (4) of human-likeness.

Figure 1. Level (1) (hook)

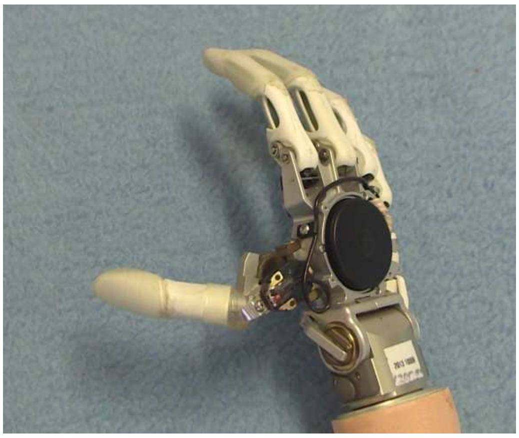

Figure 2. Level (2) (robot)

Figure 3. Level (3) (hand)

Figure 4. Level (4) (human hand)

By realizing conditions of negative and positive valence, both involving social and physical pain, in total eight different interactions with the hands were created, resulting in a total of 32 videos. Following, a list of the different conditions with negative and positive valence:

Negative conditions:

- Refusing the hand a piece of chocolate
- Refusing a handshake
- Hitting the hand
- Stinging the hand with a needle

Positive conditions:

- Handing over a piece of chocolate
- Performing a handshake
- Caressing the hand
- Touching the hand with a q-tip

A constant in the variables was the distinct intention of the human of either treating the partner hand in a friendly or in a hostile way. In order to keep environmental influences on the viewers as small as possible, all videos were recorded using the same light-blue background, the same human hand as interaction partner to the other hands and the same lightning.

2.4 Experimental procedure

The online survey was realized via SoSci Survey (Leiner, 2016) and provided to the participants on www.soscisurvey.de. The participants were taking part in the study after having signed the informed consent and did participate individually on their personal computers or mobile devices with a mean of 36 minutes to finish the survey. The participants were asked to fill in four psychometric questionnaires, two before and two after the stimulus material was presented. The stimulus material consisted of 32 short videos, which the participants were instructed to watch. The video clips were arranged in a pseudo-randomization with four different arrays and were presented without sound. After each video sequence, the participants were instructed to answer to two self-constructed items, spontaneously and without thinking too much. These two items were specifically designed for measuring cognitive and affective aspects of empathic reactions to the stimulus material.

2.5 Dependent variables

The variables of interest were self-reported measures in cognitive empathy and affective empathy. Two self-constructed items for measuring empathy in the form of a 7-point Likert scale were presented after each video. For positive conditions, response options ranged from "not pleasant at all" to "very pleasant" and for negative conditions from "not unpleasant at all" to "very unpleasant". Cited below, the text of the self-constructed items. *Item (1)* addresses an other-directed-experience (positive/negative) and aims to measure the empathic reaction within the concept of cognitive empathy. *Item (2)* addresses a self-directed-experience (positive/negative) and aims to measure the empathic reaction within the concept of affective empathy.

(1) How pleasant/unpleasant was the action for the owner of the hand in the right?

(2) How pleasant/unpleasant did the action on the hand on the right feel for you personally?

2.6 Measures

2.6.1 Toronto-Alexithymie-Skala-26 (TAS-26; Kupfer, Brosig, & Brähler, 2001)

In this study the German version, a translation from the original version of the Toronto-Alexithymia Scale (TAS) (Taylor, Ryan & Bagby, 1985; Taylor, Bagby, Ryan & Parker, 1990) was deployed. It consists of 26 items with a five-point Likert-scale. The questionnaire is a subjective measurement for dimensions of the psychological constructs of alexithymia, three sub-scales are assessing "*Schwierigkeiten bei der Identifikation von Gefühlen*" (difficulties identifying feelings), "*Schwierigkeiten bei der Beschreibung von Gefühlen*" (difficulties describing feelings) and "*extern orientierter Denkstil*" (extern oriented thinking). In short, high measures in the three scales of the TAS-26 are indicating problems of interpreting own emotional states and communicating them in an appropriate way and interpersonal difficulties. The German version of the TAS-26 holds a Cronbach's Alpha coefficient for internal consistency with values between .67 and .84 over the three scales (Kupfer, Brosig, & Brähler, 2000).

2.6.2 Individual differences in anthropomorphism questionnaire (IDAQ, Waytz, Cacioppo, & Epley, 2014)

The IDAQ is a psychometric tool for systematically measuring individual differences in anthropomorphic tendencies as a willingness to attribute personal qualities to non-human entities, which are usually considered to be unique to human nature. The questionnaire consists of 15 items containing three classes of commonly anthromorphized agents. The questionnaire measures an individual tendency to anthropomorphize by asking the subjects to evaluate animals, natural entities and technology according to the following attributes: "mind", "free will", "intentions", "consciousness", "emotions", "active", "lethargic", "good looking", "durable", "useful". Additional, fifteen non-anthropomorphic items are included for a better operationalization of anthropomorphic thinking. A German translation of the Questionnaire was devised directly by the developers of the questionnaire.

2.6.3 The Autism-Spectrum Quotient (AQ-k; Freitag et al., 2007)

The short version of the Autism-Spectrum Quotient in German language was designed to measure adults with normal intelligence and their tendency of showing traits that are associated with the autistic spectrum. The questionnaire consists of three subscales with 33 items: "*Soziale Interaktion und Spontaneität*" (social interaction and spontaneity), "*Fantasie und Vorstellungsvermögen*" (imagination and creativity) and "*Kommunikation und Reziprozität*" (communication and reciprocity). According to Freitag et al. (2007) retest-reliability and external validity of the questionnaire are satisfactory, Cronbach's alpha coefficient of internal consistency of the three scales is ranged between .65 und .87.

2.6.4 Saarbrücker Persönlichkeitsfragebogen SPF (IRI/SPF; Paulus, 2009)

The Interpersonal Reactivity Index was developed by Davis (1983) to measure empathy with regard to the multidimensional modality. The SPF (Paulus, 2009) is as a short German version of the IRI with a slight reworked factorial structure and eliminated negative formulated items. The questionnaire is a self-report measure with four subscales, representing different aspects of empathy: "*Perspektivübernahme*" (perspective taking) is defined as the tendency to spontaneously adopt the view of other people, "*Fantasie*" (fantasy) stands for the tendency of transposing oneself into the feeling and actions of fictional characters in literature or art. "*Empathische Anteilnahme*" (empathic concern) measures feelings for other people, also often labeled as "other-oriented" feelings and **"***Emotionaler Distress***"** (personal distress) measures feelings of negative tension when in socially challenging situations. Due to the multi-factorial concept of empathy, no total score is intended with the subscales. According to the author of the German version (Paulus, 2009), values for internal consistency within the subscales range between .66 und .74.

3 Results

3.1 Missing Values

The statistical analysis was carried out using IBM SPSS statistics software (version 20.0). Missing values in the scales of the questionnaires were replaced with the mean of the non-missing values according to the test-instructions. In the self-constructed items measuring cognitive and affective empathy, missing values were accepted within a maximum amount of 12% missings. The hypotheses were tested two-sided, the significance threshold was set to the *p*-value of < 0.05. For the report of effect sizes, partial Eta-squared (η^2) was calculated. The Greenhouse-Geisser adjustment was used to correct for violations of the sphericity assumption, the degrees of freedom were reported according to the correction.

3.2 Cognitive empathy and human-likeness

The independent variable "human-likeness" included four levels: hook (1), robot (2), hand (3), human hand (4), the variable "valence" had two levels, separated in negative and positive conditions. Descriptive statistics of cognitive empathy for each level of human-likeness are presented in table 1 for negative valence and in table 2 for positive valence.

Table 1

Measures of cognitive Empathy within negative Conditions

Human-likeness	*M*	*SD*	*SEM*	*95% CI*	*N*
(1) hook	2.80	1.63	0.23	[2.34, 3.25]	52
(2) robot	2.98	1.72	0.24	[2.48, 3.38]	52
(3) hand	3.65	1.70	0.24	[3.18, 4.13]	52
(4) human hand	4.33	1.50	0.21	[3.92, 4.75]	52

Note. M=means, SD=standard deviation, SEM=standard error of the means, CI=confidence intervals

Table 2

Measures of cognitive Empathy within positive Conditions

Human-likeness	*M*	*SD*	*SEM*	*95% CI*	*N*
(1) hook	2.93	1.607	0.22	[2.48, 3.38]	52
(2) robot	3.30	1.648	0.23	[2.84, 3.75]	52
(3) hand	3.40	1.612	0.22	[2.96, 3.85]	52
(4) human hand	4.90	1.232	0.17	[4.56, 5.24]	52

Note. M=means, SD=standard deviation, SEM=standard error of the means, CI=confidence intervals

The mean empathy scores and error bars with a 95% confidence interval for each condition and for each level of human-likeness are illustrated in Figure 5.

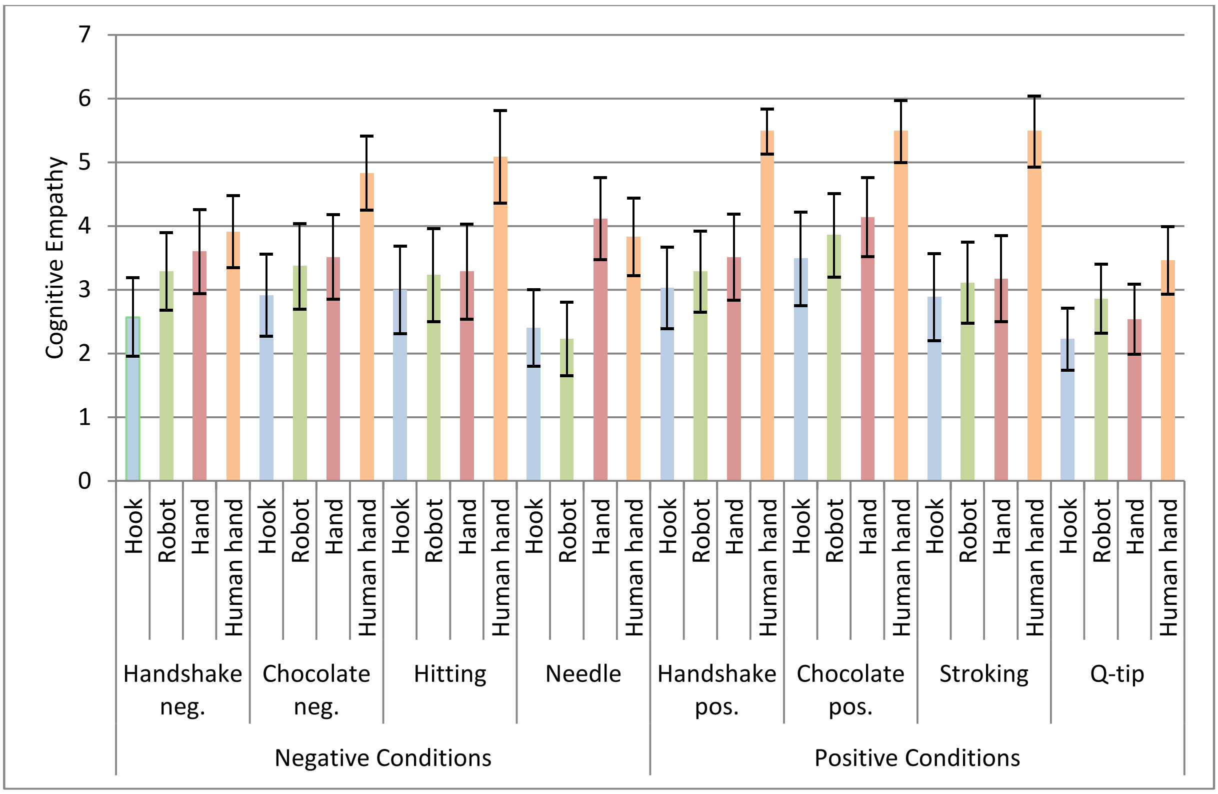

Figure 5. Mean scores of measures in cognitive empathy with a 95% confidence interval.

To test hypothesis 1.1., a two-way repeated-measures ANOVA design with the within-subjects "human-likeness" and "valence" was conducted. All effects are reported as significant at p < .05. A significant linear trend did show, F (1, 51) = 93.34, $p < .001$, $\eta 2$ = .873, indicating that with increasing human-likeness of the different hands, the empathic reaction also increased.

There was a significant *main effect of the factor "human-likeness"*, F (3, 153) = 58.46, $p < .001$, $\eta 2$ = .534 on cognitive empathy. To investigate the reported main effect, an analysis of repeated contrasts was performed, confirming that the empathy results for each level of human-likeness were significantly higher than for the level before. The empathy scores of the hand with level 2 (robot) were higher than the hand with level 1 (hook), F (1, 51) = 10,23, p = 0.002, $\eta 2$ = .167, level 3 (hand) was higher than level 2 (robot), F (1, 51) = 26,16, $p < 0.001$, $\eta 2$ = .339 and level 4 (human hand) was higher than level 3 (hand), F (1, 51) = 39.75, $p <$ 0.001, $\eta 2$ = .438.

Post-hoc analysis for the main effect of "human-likeness", applying the Bonferroni correction, revealed significant differences between each level of the hands. Mean empathy scores in level 2 was significantly higher than in level 1, *MD* = 0.28, *CI* [0.04, 0.52], p = .014, in level 3 higher than in level 2, *MD* = 0.39, CI [0.18, 0.60], p < .001 and in level 4 higher than in level 3 *MD* = 1.09, *CI* [0.61, 1.56], p < .00).There was also a significant *main effect of the factor "valence"*, F (1, 51) = 3.99, p = .051, η^2 = .073, indicating that positive and negative stimuli did lead to different empathic scores. Conditions of positive treatment, M = 3.63, *SD* = 1.37, *CI* [3.25, 4.01] significantly increased empathic reactions compared to conditions of negative valence, M = 3.44, *SD* = 1.43, *CI* [3.04, 3.84].

A significant *interaction effect* between the two main effects "human-likeness" and "valence" was found, F (2.88,146.94) = 7.03, p = <.001, η^2 = .121, indicating that the valence of the stimulus did contribute to the factor of human-likeness with different effects for the empathy ratings. Post-hoc analysis revealed significant differences between the levels of human-likeness. In the *negative condition* level 3 was significantly higher than level 1, *MD* = 0.86, *CI* [0.43, 1.29], p < .001 and level 3 was higher than level 2, *MD* = 0.67, *CI* [0.32, 1.03], p < .001. Level 4 was significantly higher than all other levels (1, 2, 3) with level 4 higher than level 3, *MD* = 0.68, *CI* [0.15, 1.21], p =.006, level 4 higher than level 2 *MD* = 1.35, CI

[0.75, 1.95], *p* < .001 and level 4 higher than level 1 *MD* = 1.54, *CI* [0.91, 2.17], *p* < .001. In the *positive condition*, level 3 was significantly higher than level 1, *MD* = 0.48, *CI* [0.17, 0,79], *p* =.009 and level 3 was higher than level 2 *MD* = 0.11, *CI* [0.34, 0.13], *p* <.001. Level 4 was significantly higher than all other levels (1, 2, 3) with level 4 higher than level 3 *MD* = 1.46, *CI* [0.96, 2.03], *p* < .001, level 4 higher than level 2, *MD* = 1.60, *CI* [1.09, 2.12], *p* < .001 and level 4 higher than level 1 *MD* = 1.97, *CI* [1.43, 2.51], *p* < .001.

The described interaction of the main effects human-likeness and valence is illustrated in Figure 6.

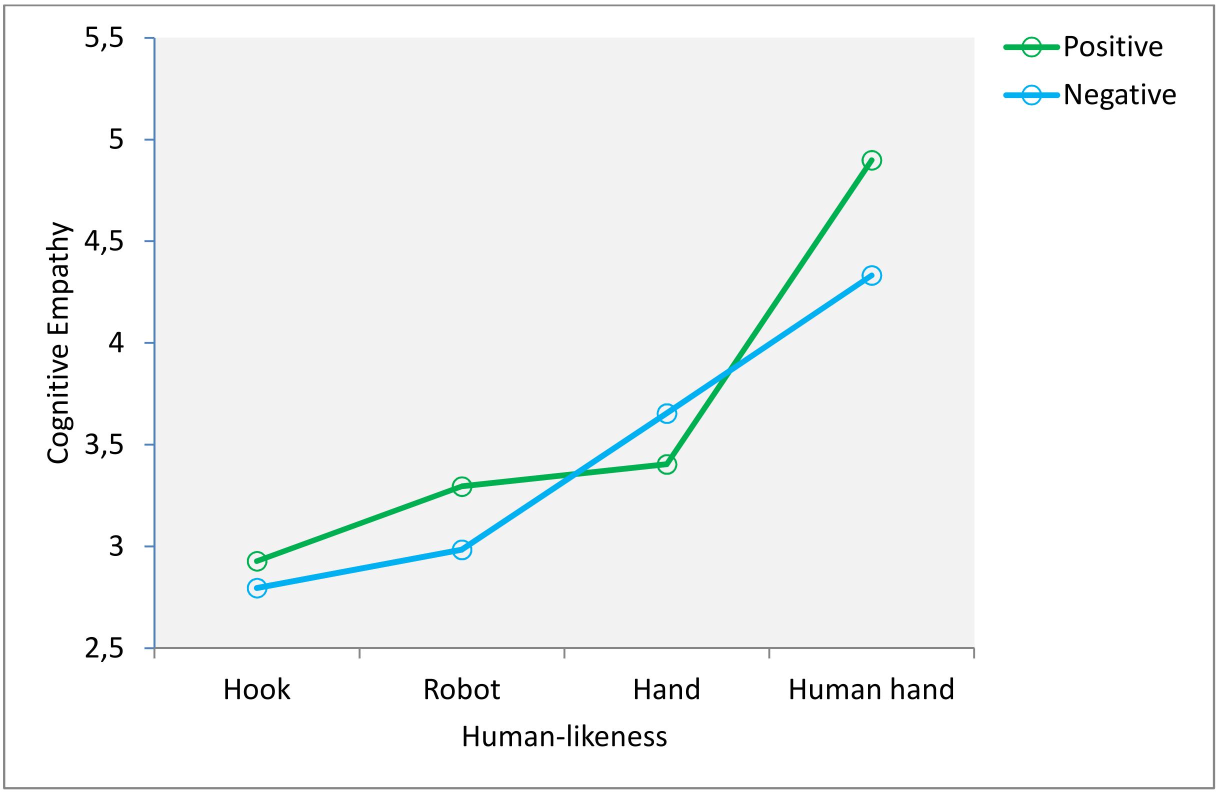

Figure 6. Interaction graph for measures of cognitive empathy measures with positive and negative valence across four levels of human-likeness.

Note. Data are means

3.3 Affective empathy and human-likeness

Descriptive statistics of affective empathy depending on the stages of human-likeness are presented for negative valence in table 3 and for positive valence in table 4.

Table 3

Measures of affective Empathy within negative Conditions

Human-likeness	*M*	*SD*	*SEM*	*95% CI*	*N*
(1) hook	3.53	1.48	0.21	[3.11,3.94]	52
(2) robot	3.89	1.47	0.20	[3.49,4.31]	52
(3) hand	4.23	1.44	0.20	[3.83,4.63]	52
(4) human hand	4.43	1.27	0.18	[4.08,4.79]	52

Note. M=means, SD=standard deviation, SEM=standard error of the means, CI=confidence intervals

Table 4

Measures of affective Empathy within positive Conditions

Human-likeness	*M*	*SD*	*SEM*	*95% CI*	*N*
(1) hook	3.50	1.43	0.20	[3.10,3.90]	52
(2) robot	4.02	1.52	0.21	[3.49,4.32]	52
(3) hand	4.07	1.36	0.19	[3.69,4.45]	52
(4) human hand	4.78	1.25	0.17	[4.43,5.12]	52

Note. M=means, SD=standard deviation, SEM=standard error of the means, CI=confidence intervals

The mean affective empathy scores and error bars with a 95% confidence interval for each condition and for each level of human-likeness are illustrated in Figure 7.

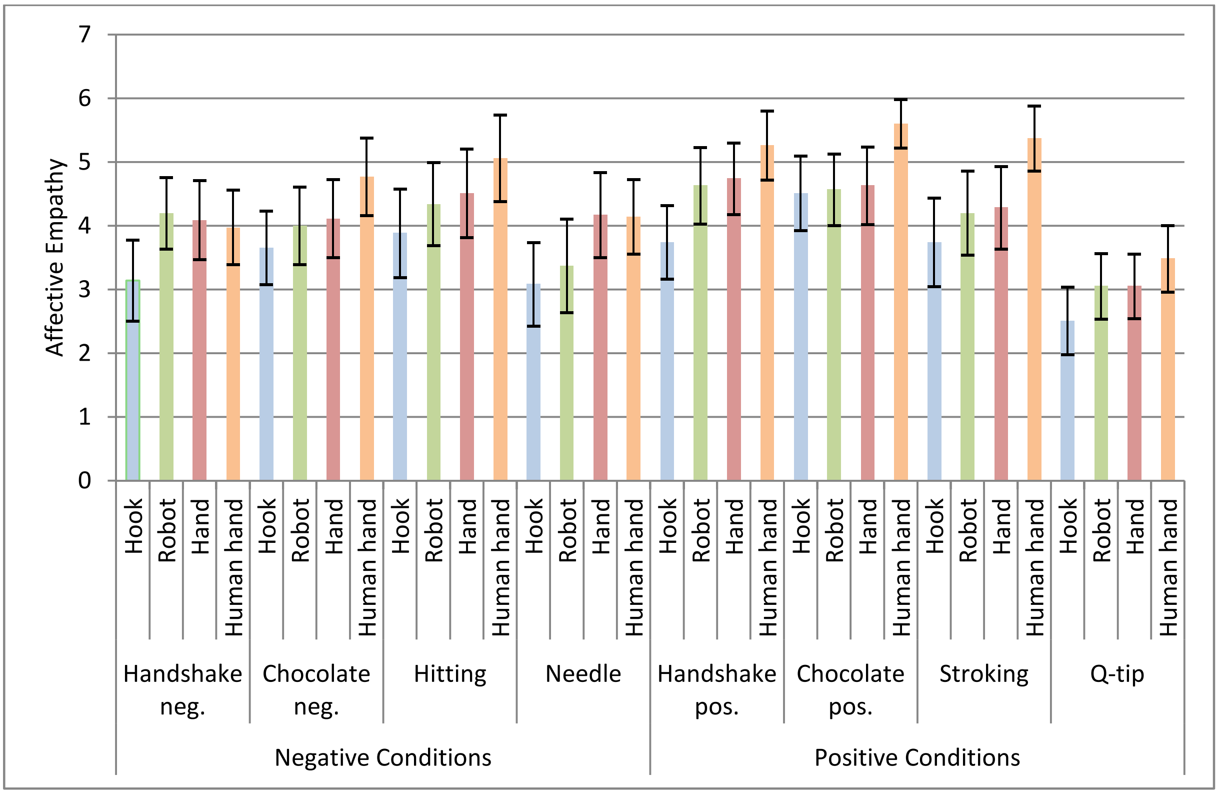

Figure 7. Mean scores of measures in affective empathy with a 95% confidence interval.

To test hypothesis 1.2, a two-way repeated-measures ANOVA design was conducted. All effects are reported as significant at $p < .05$. Post-hoc analyses applying the Bonferroni correction was performed to explore differences between group means of affective empathy. The Greenhouse-Geisser adjustment was used to correct for violations of the sphericity assumption with the interaction effect of human-likeness and valence, the degrees of freedom and p-values were reported according to the correction.

The two-way ANOVA showed a significant linear trend, F (1, 51) = 45.859, $p < .001$, $\eta 2 = .473$, indicating that with increasing human-likeness of the different hands, the empathic reaction also increased. The analysis also revealed a significant *main effect of the factor "human-likeness"*, F (3, 153) = 24.55, $p < .001$, $\eta 2 = .325$. Post-hoc analysis applying the Bonferroni correction revealed significant differences between each hand, except level 2 (robot) and level 3 (hand), $p = .33$. Mean empathy scores in level 2 were significantly higher than in level 1, *MD* = 0.45, *CI* [0.17, 0.72], $p < .001$, in level 3 higher than in level 1, *MD* = 0.64, *CI* [0.32, 0.95], $p < .001$ and in level 4 higher than in level 3 *MD* = 0.45, *CI* [0.08, 0.82], $p = .009$.

The factor "valence" had no significant main effect on the ratings in affective empathy.

An *interaction effect* between the two factors "human-likeness" and "valence" was found, F (2.821, 143.846) = 1.28, p = .036, η^2 = .055, indicating that the valence of the stimulus did significantly contribute to the factor of human-likeness with different effects for the empathy ratings. Post-hoc analysis revealed significant differences between some of the levels of human-likeness.

In the *negative condition* level 3 was significantly higher than level 1, *MD* = 0.70, *CI* [0.272, 1.136], p < .001, level 4 was higher than level 1, *MD* = 0.905, *CI* [0.391, 1.420], p < .001 and level 4 was higher than level 2 *MD* = 0.534, *CI* [0.017, 1.050], p =.039.

In the *positive condition*, level 2 was significantly higher than level 1, *MD* = 0.522, *CI* [0.174, 0.871], p =.001 and level 3 was higher than level 1 *MD* = 0.571, *CI* [0.226, 0.915], p < .001. Level 4 was significantly higher than all other levels (1, 2, 3) with level 4 higher than level 3 *MD* = 0.704, *CI* [0.229, 1.178], p = 001, level 4 higher than level 2 *MD* = 0.752, *CI* [0.321, 1.183], p < .001 and level 4 higher than level 1 *MD* = 1.274, *CI* [0.797, 1.751], p < .001.

The described interaction of the main effects of "human-likeness" and "valence" is illustrated in Figure 8.

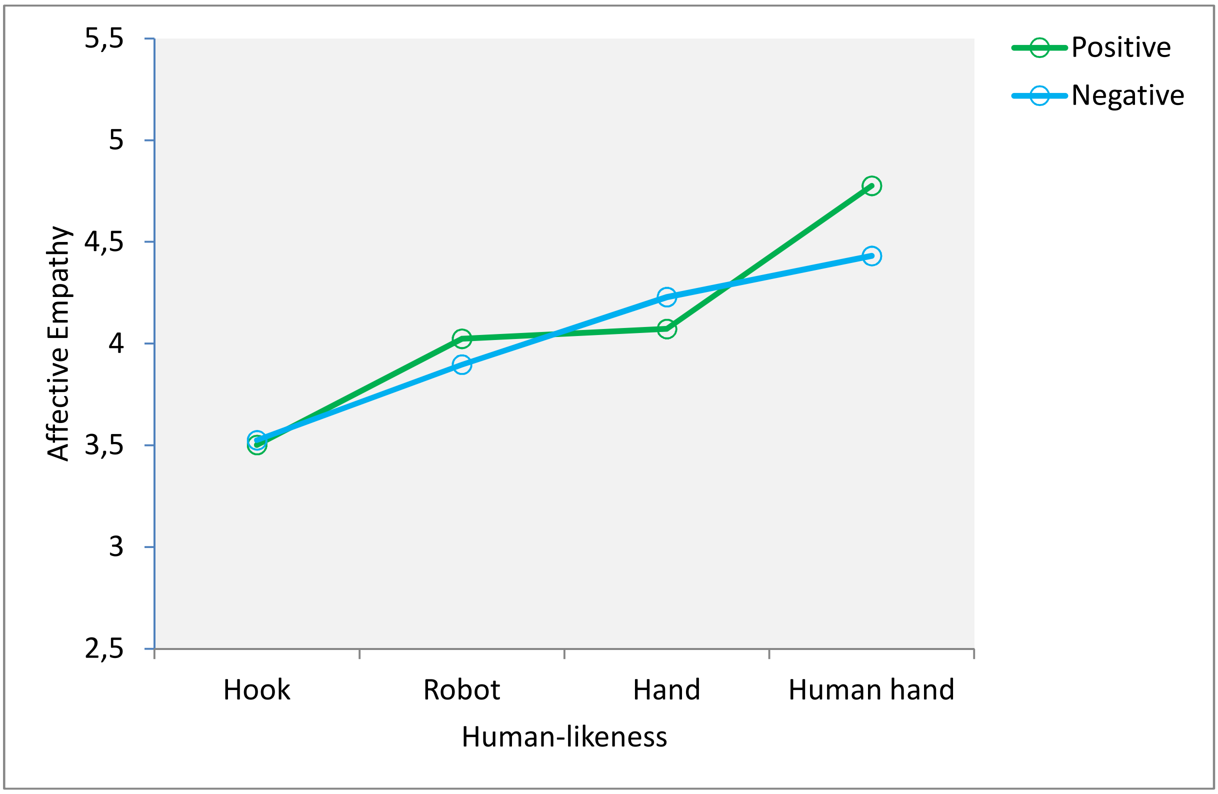

Figure 8. Interaction graph for measures of affective empathy measures with positive and negative valence across four levels of human-likeness.

Note. Data are means.

3.4 Personal traits, sociodemographic characteristics and empathy

Personal traits of the participants were assessed through psychometric questionnaires as described in 2.6., sociodemographic data of the participants is described in paragraph 2.2. To test hypotheses 2 and 3, a Pearson's Product-Moment Correlation was used to determine the relationship between self-reported empathy, personal traits and differences in individual sociodemographic data. Due to the small number of empiric evidence about individual differences in empathic reactions towards robots, correlation analysis was conducted with all subscales and total scores of the deployed psychometric questionnaires and the sociodemographic data for both affective and cognitive empathy. Variables of sociodemographic data were age, gender, actually living or having lived together with a pet, owning a technical device with a name given and the highest level of education.

Contrary to the expectations, this study did not find a significant correlation between individual differences in personal traits on self-reported measures of cognitive or affective empathy. There was also no correlation of sociodemographic data with scores in cognitive or affective empathy.

4 Discussion

Based on the current discourse about human-robot-interaction, the ways robots are emotionally perceived according to their outer appearance, this work was conducted to evaluate empathic reactions towards different robotic entities on an anthropomorphic spectrum. One general challenge of empirically researching anthropomorphic tendencies in individuals is, among others issues, socially desirable responding. Attributing mental states to human or non-human entities or reporting emotional reactions towards an inanimate object might be an attitude people are reluctant to display overtly. When developing a scale for measuring anthropomorphism, Chin et al. (2005) experienced that people showed less hesitation in self-reporting anthropomorphism towards socially "appropriate" targets of humanization, like e.g. pets, than towards inanimate or technical objects like cars or computers. In this study, the effect of reluctance in reporting potential improper emotional reactions towards robots did not show. Accordingly, the findings of the current study provide support for the hypothesis, that people experience an increased degree of empathy with robots of higher humanlike appearance than with more mechanical looking ones. The effect of a higher empathic reaction according to human-likeness occurred on both, a cognitive and an affective level.

The effect of showing affective empathic reactions for technical agents are compatible with previous studies, revealing that humans experience empathy for robots in abusive situations (Bartneck & Hu, 2008; Hoenen et al., 2016). Vaes et al. (2016) observed similar empathic reactions to painful situations of robots and humans in bottom-up processes, but reported difficulties of the subjects in taking the perspective of a robot. Contrary to this, subjects in the present study did clearly report cognitive empathy for the robotic agents. Still, this result should be considered with caution, as a single item of self-reported data might not satisfy the complex phenomenon of empathy. Moreover, mentally relating to artificial creatures on higher cognitive processes is quite new territory for humans anyways. Still, this result is in line with findings, that people show higher tendencies of taking the perspective of and ascribing a mind to inanimate entities, which represent more humanlike appearance (Gray et al., 2007; Krach et al., 2008; Riek et al., 2009).

Regarding the influence of friendly or hostile interaction with the robots, the findings of the current study differ from previous research. Measuring empathic reactions while watching interactions of a human with a human and a robot, Rosenthal-von der Pütten, Schulte, Eimler, Sobieraj, Hoffmann, Maderwald, et al. (2014) found no difference in the empathic reaction for humans and robots in the positive condition, while hostile treatment did lead to significant differences between HHI and HRI, with a higher negative empathic concern for the human "victim". Contrary to the findings of Rosenthal-von der Pütten et al. (2014), in this study, significant differences in empathizing with robotic hands were found in both, positive and negative treatment. Furthermore, on an overall level, the empathic measures found in this study were higher for positive conditions of the robots than for negative ones, both in cognitive and affective empathy. However, these measures are in contrast to common results of social interactions with positive and negative treatments, as hostile or painful conditions usually lead to higher empathic reactions than friendly ones. There are several possible explanations for this discrepancy. For one, it might be an effect of the stimuli themselves, as the portrayal of positive interactions with the partner hands were probably perceived as more precise than the ones of negative treatment. To give an example, friendly versus hostile treatment was implemented by either performing a handshake with the robotic hand or refusing the handshake. In the short time sequences of the video clips, shaking a hand might be a less unambiguous act than pulling a hand away from a handshake and, by that, the latter might have lead to weaker effects within the empathic reactions. Influences of the individual interpretation of the deployed stimuli objects might also have contributed to the unusual effect of higher empathy within positive conditions. For example, one participant did not clearly identify the stinging needle as an aversive item, but suspected it to be a technical tool to supply the robotic hand with functional liquids. Following this participants' report, the deployment of some of the objects might have caused contentual ambiguity.

Another potential reason for the unusual effect of valence in measures of cognitive empathy could be a question of statistics. With the number of participants (N = 52) it could be argued that this sample size might represent a lack of power for showing the same effects in both, negative and positive conditions.

Although with no statistical significance, for the most humanlike robotic hand (level 3) an inversion of the influence of valence to the empathic measures appeared. This prosthetic hand with the skin-like cover and a highly android look was clearly the most ambiguous of the robotic hands. With no precise evidence for the inverted effect, it can only be hypothesized that the ambiguous look of the hand made it harder for the participants to categorize and therefore might have lead to a stronger orientation towards the stimuli themselves. This means, that the strong anthropomorphic physiognomy might have disarrayed the empathic assessment of the participants in a way that the valence of the stimuli was of more presence and therefore did lead to the result of higher empathy with negative stimuli, as it did for the rest of the hands.

Contrary to the expectations, this study did not find evidence for a significant influence of differences in personality traits and sociodemographic data on measures of empathy. There is no specific explanation for this result but it seems possible that these results may be attributed to the peculiarity of the participating individuals. It should be noted that the composition of the participants in this study did clearly deviate from the majority of participants in empiric studies. Usually, in empirical studies mostly students participate, which represent a rather homogenous group of people regarding age and education. With a medium age of 42 years and a range from 25 to 69 years, 63.5% participants holding a university degree and 78.8% being employed or self-employed, subjects in this study clearly differ in age, education and profession from the usual cohort of younger students. Furthermore, the people participating are mainly socialized in an environment of creative industry and humanities and therefore have to be considered as a specific segment of the general population, which is not fully representative.

4.1 Limitations

The limitations which have to be considered in this work are mostly attached to the problem of creating appropriate stimulus material for the experimental part of the study. Most empiric studies in the field of robotics face problems with finding access to different robotic entities with adequate anthropomorphic features that differ only in the factor of human-likeness. Systematically varying this factor and keeping all other influencing variables of the robot constant poses a huge challenge in the experimental design.

In this study, moving prosthetic hands were used to create the impression for the participants that they deal with a fully embodied robotic creature. As the self-reported measures of empathy were aimed to evaluate an artificial being, the verbal formulation of the items related to the "owner" of the hand as a complete entity. By using the possibility to make comments at the end of the survey, some participants reported that with continuously watching the stimulus material, they started to reflect on what they were observing. One participant reported that she was starting to think about the developer of the robots and whether it would be a problem for him or her if they were failing to complete some of the tasks. Another participant reported that he did not perceive the hands as a part of a robot, but as prosthetic hands and was thinking about the potential human owner. In these cases, the basic concept of the study, to evaluate reactions without intense contemplating about the scenery, was not fulfilled, with potential consequences for the evaluated results. It can only be reasoned, that if an increased cognitive activation of the participants during the trial took place, this might especially be in conflict with fast, automatic responses, as designated for measuring the affective component of the empathic reactions.

Furthermore, with a total of 32 videos to watch, one participant reported that he had growing difficulties in responding adequately to each video, as he got puzzled by the similarity of them. All in all, the rather long overall process time of the complete survey with an average of 38 minutes can be regarded as the main factor for participants to drop out during the process of answering. For future research a more reduced study design with less questionnaires or video stimuli would probably lead to a higher number of subjects finishing and thus to a larger sample size.

5 Conclusion

At the present time, mentally relating to artificial emotions is still a quite new territory in human psychology. Empiric research, discussing empathy in human-robot-relationships from the human perspective is still scarce and only a few studies have been able to conduct systematic research on this topic. The findings in this study contribute to the present discussion about human capacities and dispositions of responding to non-human entities and on the subsequent potential to establish social and emotional relationships with inanimate objects. Yet, merely on the basis of self-reported measures, the findings reported in this work are consistent with previous evidence that people are capable of relating emotionally to a robotic agent. A significant finding that emerged from this study is that humanlike design and morphology of robotic entities has an important impact on the ability to empathize with robots The results in the study are in consensus with recent evidence in literature and research of varying grades of empathy towards robots depending on their anthropomorphic appearance.

6 References

Adolphs, R. (2005). Could a robot have emotions? Theoretical Perspectives from Social Cognitive Neuroscience. In J.M. Fellous & M.A. Arbib (Eds.), *Who needs emotions*. Oxford: University Press.

Asada, M. (2015). Towards artificial empathy. *International Journal of Social Robotics, 7*(1), 19-33.

Baron-Cohen, S. Wheelwright, S., Skinner, R., Martin, J., Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders, 31*(1).

Bartneck, C. & Hu, J. (2008).Exploring the Abuse of Robots. *Interaction Studies*, *9*(3), 415-433.

Bartneck, C., Kulic, D., & Croft, E. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social·Robotics, 2(1),* 71-81. doi: 10.1007/s12369-008-0001-3

Bartneck, C., van der Hoek, M., Mubin, O., & Mahmud, A. A. (2007*). "Daisy, daisy, give me your answer do!" - Switching off a robot.* Proceedings of the 2nd ACM/IEEE International Conference on Human-Robot Interaction, Washington DC, 217 - 222. doi:10.1145/1228716.1228746

Batson, C.D., Lishner, D.A., Cook, J., & Sawyer, S (2005) Similarity and nurturance: two possible sources of empathy for strangers. *Basic and Applied Social Psychology, 27*(1), 15-25.

BBC, (2016). Sad, but the legacy lives on [Video]. Retrieved 6th September 2017, from http://www.bbc.com/news/av/science-environment-37521318/rosetta-mission-really-sad-but-the-legacy-lives-on

Bisio, A., Sciutti, A., Nori, F., Metta, G., Fadiga, L., Sandini, G., et al. (2014). Motor contagion during human-human and human-robot interaction. *PLoS One, 9*(8). *doi*: 10.1371/journal.pone.0106172

Bouquet, C.A., Shipley, T.F., Capa, L.R., & Marshall, P.J. (2011). Motor contagion: goal-directed actions are more contagious than non-goal directed actions. *Experimental Psychology, 58* (1), 71-80.

Breazeal, C. (1998). A motivational system for regulating human-robot interaction, in *Proceedings of the 15th National Conference on Artificial Intelligence (AAAI98*), WI, 54-66.

Breazeal, C. (2002). *Designing sociable robots*. Cambridge, USA: MIT Press.

Breazeal, C. (2003). Social interactions in HRI: The robot view. *IEEE Transaction on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 34*(2), 181-186. doi:10.1109/TSMCC.2004.826268

Breazeal, C., & Brooks, R. (2005). Robot emotion: A functional perspective. In J.M. & M.A. Arbib (Eds.), *Who Needs Emotions*. Oxford: University Press.

Čapek, K. (1920/2004). *R.U.R. (Rossum's Universal Robots).* London, New York: Penguin Books. ISBN: 978-0141182087

Caporael, L. R., & Heyes, C. (1997). Why anthropomorphize? Folk psychology and other stories. In R.W. Mitchell, N.S. Thompson & H.L. Miles (Eds.), *Anthropomorphism, Anecdotes, and Animals* (pp. 59-73). Albany, NY: University of New York Press.

Carpenter, J. (2016). *Culture and Human-Robot Interaction in Militarized Spaces.* London and New York: Routledge, Taylor & Francis Group.

Chan, A.A.Y.H. (2013). Anthropomorphism as a conservation tool. *Biodeversity and Conversation 22*(8), 1577-1589.

Chin, M.G., Ryan, E.Y., Clark, B.R., Ballion, T., Dolezal, M.J., Shumaker, R., & Finkelstein, N. (2005). Developing and anthropomorphic tendencies scale. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 49*(13), 1266-1268.

CNN, (2015). Is it cruel to kick a robot dog? [Video].Retrieved 22th August, 2017, from: http://edition.cnn.com/2015/02/13/tech/spot-robot-dog-google/index.html

Cullen, H., Kanai, R., Bahrami, B, & Rees, G. (2014). Individual differences in anthropomorphic attributions and human brain structure. *Social Cognitive and Affective Neuroscience*, *9*(9), 1276–1280.

Cruz, J., & Gordon, R. M. (2002). Simulation theory. In L. Nagel (Eds.), *Encyclopedia of Cognitive Science*. Macmillan.

Damasio, A. (1999). *The feeling of what happens: Body, emotion and the making of consciousness*. New York: Harcourt, Brace, and Co.

Damiano, L., Dumouchel, P., & Lehmann, H. (2015). Artificial empathy: An interdisciplinary investigation. International Journal of Social Robotics, 7(1), 3-5.

Darling, K. (2017). The future of human-robot interaction [Presentation]. Aspen Ideas Festival 2017. Retrieved 7th January, 2018, from: https://www.aspenideas.org/speaker/kate-darling

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. Journal of Personality and Social Psychology, 44(1), 113-126.

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews, 3*(2), 71-100.

Decety, J., Lamm, C. (2006). Human empathy through the lens of social neuroscience. *Scientific World Journal, 6*, 1146-1163.

Decety, J., & Lamm, C. (2009). Empathy versus personal distress: recent evidence from social neuroscience. In J. Decety & W. Ickes (Eds.), *The Social Neuroscience of Empathy.* Cambridge: MIT Press.

Decety, J., & Sommerville, J. A. (2003). Shared representations between self and others: A social cognitive neuroscience view.*Trends in Cognitive Science*, *7(12)*, 527-533.

De Graaf, M.M.A. (2016). An ethical evaluation of human–robot relationships. *International Journal of Social Robotics, 8(*4), 589-598. doi: 10.1007/s12369-016-0368-5

De Waal, F. B. M. (1999). Anthropomorphism and anthropodenial: Consistency in our thinking about humans and other animals. *Philosophical Topics*, *27*(1), 255–280.

Dovidio, J.F., Johnson, J.D., Gaertner, S.L., Pearson, A.R., Saguy, T., & Ashburn-Nardo, L. (2010). Empathy and intergroup relations. In M. Mikulincer & P.R. Shaver (Eds.), *Prosocial Motives, Emotions, and Behavior: The better Angels of our Nature*. Washington, DC: American Psychological Association.

Eisenberger, N.I., Lieberman, M.D., & Williams, K.D. (2003). Does rejection hurt? An fMRI study of social exlclusion. *Science, 302*(5643), 290-292. doi: 10.1126/science.1089134

Ekman, P., & Friesen, W.V. (1978). *Manual for the facial action coding system*. Palo Alto: Consulting Psychologists Press.

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

Euron, European Robotics Research Network. (2006). Roboethics Roadmap Book. Retrieved 12th December, 2017 from: http://www.roboethics.org/atelier2006/docs/ ROBOETHICS%20ROADMAP%20Rel2.1.1.pdf

Europäisches Parlament. (2015). Entwurf eines Berichts mit Empfehlungen an die Kommission zu zivilrechtlichen Regelungen im Bereich Robotik, (2015/2103(INL). Retrieved 12th January, 2017, from http://www.europarl.europa.eu/oeil/popups/ ficheprocedure.do?lang=en&reference=2015/2103(INL)#documentGateway

Fang-Wu, T. (2016). Child Perception of humanoid robot appearance and behavior. *International Journal of Human-Computer Interaction*, *32*(6), 493-502.

Ferrari, F., Paladino, M.P., & Jetten, J. (2016). Blurring human–machine distinctions: Anthropomorphic appearance in social robots as a threat to human distinctiveness. *International Journal of Social Robotics, 8*(2), 287-302.

Fisher, J.A. (1991). Disambiguating anthropomorphism: An interdisciplinary review. *Perspectives in Ethology,9*, 49-85.

Freitag, C. M., Retz-Junginger, P., Retz, W., Seitz, C., Palmason, H., Meyer, J., …. von Gontard, A. (2007). Evaluation der deutschen Version des Autismus-Spektrum-Quotienten (AQ) - die Kurzversion AQ-k. *Zeitschrift für Klinische Psychologie und Psychotherapie, 36*(4), 280–289.

Freud, S. (1919). *Das Unheimliche*. In Studienausgabe, Bd. IV. Frankfurt a.M. 1970: Fischer.

Fuchs, T. (2014). The virtual other: Empathy in the age of virtuality. *Journal of Consciousness Studies*, *21*(5-6), 152-173.

Gallese, V. & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Science, 2*(12), 493-501.

Gallese, V., Keysers, C., & Rizzolatti. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Science, 8*(9), 396-403.

Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007). The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *Neuroimage, 35*(4), 1674–84.

Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 67-82). Hillsdale, NJ: Erlbaum

Gray H.M., Gray, K., & Wegener, D.M. (2007). Dimensions of mind perception. *Science 315*(5812*)*, 619. doi: 10.1126/science.1134475

Gopnik, A., Meltzoff, A. (1997). *Words, thoughts, and theories*. Cambridge, Mass.: Bradford,MIT Press.

Gopnik, A., Wellman, H. (1994). The ‚Theory-Theory’. In L. Hirschfield and S. Gelman (Eds.), *Domain specificity in culture and cognition*. New York: Cambridge University Press.

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. doi: 10.1037/0022-3514.89.6.937

Hegel, F., Krach, S., Kircher, T., Wrede, B., & Sagerer, G. (2008). Understanding social robots: A user study on anthropomorphism. *In RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*, Munich. doi:10.1109/ROMAN.2008.4600728

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

Hodges, S. & Myers, M. (2007). Empathy. In R. F. Baumeister & K. D. Vohs (Eds.), *Encyclopedia of Social Psychology* (pp. 297-298). Thousand Oaks, CA: SAGE Publications Ltd.

Hoenen, M., Lübke, K.T., & Pause, B.M. (2016). Non-anthropomorphic robots as social entities on a neurophysiological level. *Computers in Human Behavior, 57*,182-186.

Höök, K. (2008).The affective Loop. In H. Oinas-Kukkonen, P. Hasle, M. Harjumaa, K. Segerstah, & P. Ohrtrom (Eds.), *PERSUASIVE 2008*. Berlin, Heidelberg: Springer Verlag.

Hudlicka, E. (2008). What are we modeling when we model emotion? *International Journal of Synthetic Emotions, 2*(1), 26-79.

Huffman, C.A. (2012). Archytas of Tarentum: *Pythagorean, philosopher and mathematician king*. Cambridge: University Press.

Hume, D., Colver, A. W., Price, J. V., & Hume, D. (1976). *The natural history of religion*. Oxford, England: Clarendon Press.

Jackson, PL., Rainville, P., & Decety, J. (2006).To what extent do we share the pain of others? Insight from the neural bases of pain empathy. *Pain, 125*(1-2), 5-9.

Jackson, PL., Brunet E., Meltzoff, AN., Decety J. (2006).Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. Neuropsychologia, *44*(5), 752-61.

Jackson, PL, Meltzoff AN, Decety J. (2005). How do we perceive the pain of others? A window into the neural processes involved in empathy. *NeuroImage, 24*(3), 771-779.

Jakob, P. & Jeannerod, M. (2005). The motor theory of social cognition: a critique. T*rends in cognitive sciences, 9*(1), 21-25. doi:10.1016/j.tics.2004.11.003

James, W. (1884). What is an emotion? *Mind, os-IX*(34), 188-205. doi: 10.1093/mind/os-IX.34.188

Kiesler, S., Powers, A., Fussell, S.R., & Torrey, C. (2008). Anthropomorphic interactions with a robot and a robot-like agent. *Social Cognition, 26*(2), 169-181.

Krach, S., Hegel, F., Wrede B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS ONE, 3*(7) e2597. doi:10.1371/journal.pone.0002597

Krämer, N.C., von der Pütten, A., & Eimler, S. (2012). Human-agent and human-robot interaction theory: Similarities to and differences from human-human interaction. In M. Zacarias, J.V. de Oliveira (Eds.), *Human-Computer Interaction: The Agency Perspective,* (pp.215-240). Berlin, Heidelberg: Springer Verlag.

Krebs, D. L. (1975). Empathy and altruism. *Journal of Personality and Social Psychology, 32*, 1134-1146.

Kupfer, J., Brosig, B. & Brähler, E. (2001). *TAS-26: Toronto-Alexithymie-Skala-26 (deutsche Version)*. Göttingen: Hogrefe.

Kupferberg, A., Huber M., Helfer, B., Lenz, C., Knoll, A., & Glasauer S. (2012). Moving just like you: Motor interference depends on similar motility of agent and observer *PLoS One, 7,* e39637. doi: 10.1371/journal.pone.0039637

Lamm, C. Batson, CD, & Decety, J. (2007). The neural substrate of human empathy: Effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience,19*(1), 42-58.

Lamm C, Bukowski, H., Silani G. (2016). From shared to distinct self– other representations in empathy: evidence from neurotypical function and socio-cognitive disorders. *Philosophical Transactions of the Royal Society, SeriesB, Biological sciences B371*. doi:10.1098/rstb.2015.0083

Lamm, C., & Majdandžić, J. (2014). The role of shared neural activations, mirror neurons, and morality in empathy – A critical comment. *Neuroscience Research, 90*, 15-24. doi:10.1016/j.neures.2014.10.008

Leiner, D. J. (2016). SoSci Survey (Version 2.6.00) [Computer software]. Available at https://www.soscisurvey.de

Letheren, K., Kuhn, K.-A. L., Lings, I., Pope, N.K. (2016). Individual difference factors related to anthropomorphic tendency. *European Journal of Marketing, 50*(5/6), 973-1002.

MacDorman, KF (2006). Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. *ICCS/CogSci-2006 long symposium: Toward social mechanisms of android science*, 26-29.

MacDonald, G., & Leary, MR. (2005). Why does social exclusion hurt? The relationship between social and physical pain. *Psychological Bulletin,131*, 202–223.

Meltzoff, A. N., & Brooks, R. (2001). "Like me" as a building block for understanding other minds: Bodily acts, attention, and intention. In B. F. Malle, L. J. Moses, & D. A.

Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition*, (pp. 171-191). Cambridge: MIT Press.

Misselhorn, C. (2009). Empathy with Inanimate Objects and the Uncanny Valley. *Minds & Machines,19,* 345-359.

Moran, M.E. (2007). The da Vinci robot. *Journal of Endourology, 20*(12), 986-990. doi:10.1089/end.2006.20.986

Mori, M. (1970). The uncanny valley. *Energy, 7*(4), 33–35.

Nass, C., Steuer, J., & Tauber, E.R. (1994).Computers are social actors. In B. Adelson, S. Dumais, & J. Olson (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp.72-78). doi:10.1145/191666.191703

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

Nishio, S., Koichi, T., Hidenobu, S., & Ishiguro, H. (2013) Teleoperated android robot as emotion regulation media. *International Journal of Social Robotics, 5,* 563-573. doi:10.1007/s12369-013-0201-3

Nitsch, V. & Popp, M. (2014). Emotions in robot psychology. *Biological Cybernetics*, *108*(5), 621-629. doi:10.1007/s00422-014-0594-6

Oberman, L.M., McCleery, J.P., Ramachandran, V.S, & Pineda, J.A. (2007). EEG evidence for mirror neuron activity during the observation of human and robot actions: Toward an analysis of the human qualities of interactive robots. *Neurocomputing, 70*, 2194-2203.

Oztop, E., Franklin, D.W., Chaminade, T., & Cheng, G. (2005).Human-humanoid interaction: Is a humanoid robot perceived as a human? *International Journal of Humanoid Robotics, 2*(4), 537-559.

Panksepp, J. & Watt, D. (2011). What is basic about basic emotions? Lasting lessons from affective neuroscience. *Emotion Review*, *3*(4), 1-10. doi:10.1177/1754073911410741

Paulus, C. (2009). *Saarbrücker Persönlichkeits-Fragebogen zu Empathie (SPF)*. Retrieved 27th September 2016 from http://bildungswissenschaften.unisaarland.de/personal/paulus/empathy/SPF.htm

Piaget, J. (1929). The child's conception of the world. *Mind, 38*(152), 506-513.

Picard, R.W. (2003). What does it mean for a computer to "have" emotions? In R. Trappl, P. Petta, S. Payr (Eds.), *Emotions in Humans and Artifacts* (pp.213-235). Cambridge:MIT Press.

Powers, A., & Kiesler, S. (2006). The advisor robot: tracing people's mental model from a robot's physical attributes. *Proceedings of the 1st Annual Conference on Human-Robot Interaction,* 218-225. doi 10.1145/1121241.1121280

Press, C., Bird, G., Flach, R., & Heyes, C. (2005). Robotic movement elicits automatic imitation. *Cognitive Brain Research, 25(3)*, 632-640. doi:10.1016/j.cogbrainres.2005.08.020

Preston, S. (2007). A perception-action model for empathy. In T. Farrow & P. Woodruff (Eds.), *Empathy in Mental Illness* (pp. 428-447). Cambridge: University Press.

Preston, SD. & De Waal, F.B.M. (2002) Empathy: Its ultimate and proximate bases. B*ehavioral and Brain Sciences*, *25*(1), 1-20. doi:10.1017/S0140525X02000018

Prinz, J.J. (2004). Embodied emotions. In R. C. Solomon (Ed.), *Thinking about feeling: Contemporary philosophers on emotions* (pp. 44–58). New York: Oxford University Press.

Reeves, B. & Nass, C. (1996).The media equation. How people treat computers, televions, and new media like real people and places. USA, New York: Cambridge University Press.

Riek, L.D., Rabinowitch, T.C., Chakrabarti, B., & Robinson, P. (2009). How anthropomorphism affects empathy toward robots. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction (HRI '09),* 245-246. doi: 10.1145/1514095.1514158

Rizzolatti, G., & Craighero, L. (2004). The mirror neuron system. *Annual Review of Neuroscience, 27,* 169-192. doi:10.1146/annurev.neuro.27.070203.144230

Rosenthal-von der Pütten, A.M., Krämer, N.C., Hoffmann, L., Sobieraj, S., Eimler, S.C. (2013). An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics, 5*(1), 17-34. doi:10.1007/s12369-012-0173-8

Rosenthal van der Pütten, A., Schulte, F.P., Eimler, S.C., Sobieraj, S., Hoffmann, L., Maderwald, S., … Krämer, N. (2014).Investigations on empathy towards humans and robots using fMRI. *Computers in Human Behavior, 33*, 201-212.

Root-Bernstein M, Douglas L, Smith A, & Veŕıssimo D. (2013) Anthropomorphized species as tools for conservation: utility beyond prosocial, intelligent and suffering species. *Biodiversity and Conservation, 22*(8),1577–1589. doi:10.1007/s10531-013-0494-4

Ruby, P. & Decety, J. (2004). How would you feel versus how do you think she would feel? A neuroimaging study of perspective taking with social emotions. *Journal of Cognitive Neuroscience, 16*(6), 988-999. doi: 10.1162/0898929041502661

Rütgen, M., Seidel, E-M., Silani, G., Riečanský, I., Hummer, A., Windischberger, C., Petrovic, P., & Lamm, C. (2015). Placebo analgesia and its opioidergic regulation suggest that empathy for pain is grounded in self pain. *Proceedings of the National Academy of Sciences*, 112*(*41), E5638-5646. doi:10.1073/pnas.1511269112

Şahin, E., Cakmak, M., Dogar, M.R., Ugur,E., & Göktürk, Ü. (2007).To afford or not to afford: A new formalization of affordances toward affordance-based robot control. *Adaptive Behavior, 5*(15), 447-472.

Scheele, D., Schwering, C., Elison, J.T., Spunt, R., Maier, W., & Hurlemann, R. (2015). A human tendency to anthropomorphism is enhanced by oxytocin. *European Neuropsychopharmacology, 25*(10), 1817-1823.

Schlobinski, P. & Siebold, O. (2008). *Wörterbuch der Science-Fiction*. Frankfurt: Peter Lang Verlag.

Shanton, K. & Goldman, A. (2010). Simulation Theory. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(4), 527-538. doi:10.1002/wcs.33

Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: review of literature and implications for future research. *Neuroscience & Biobehavioral Reviews, 30*(6), 855–863. doi:10.1016/j.neubiorev.2006.06.011

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R.J., & Frith C.D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science 303,* 1157-1162. doi: 10.1126/science.1093535

Suzuki, Y., Galli, L., Ikeda, A., Itakura, S., & Kitazaki M. (2015).Measuring empathy for human and robot hand pain using electroencephalography. *Scientific Reports 5:* 15924. doi: 10.1038/srep15924

Tai, Y.F., Scherfler, C., Brooks, D.J., Sawamoto, N., & Castiello, U. (2004). The human premotor cortex is 'mirror' only for biological actions. *Current Biology, 14*(2), 117-120.

Taylor, N. (2012). Cinematic Perspectives on Digital Culture. Consorting with the Machine. Hampshire: Palgrave Macmillan.

Tisseron, S., Tordo, F., & Baddoura, R. (2015). Testing Empathy with Robots. *International Journal of Social Robotics, 7*(1), 97-102. doi:10.1007/s12369-014-0268-5

Urquiza-Haas, E.G., & Kotrschal, K. (2015). The mind behind anthropomorphic thinking: attribution of mental states to other species. *Animal Behaviour, 109*, 167-176. doi:10.1016/j.anbehav.2015.08.011

Vaes, J., Meconi, F., Sessa, P., & Olechowski, M. (2016).Minimal humanity cues induce neural empathic reactions towards non-human entities. *Neuropsychologia, 89*, 132-140. doi: 10.1016/j.neuropsychologia.2016.06.004

Vytal, K., & Hamman, S. (2010). Neuroimaging support for discrete neural correlates of basic emotions: a voxel-based meta-analysis. *Journal of Cognitive Neuroscience, 22*(12), 2862-2865. doi: 10.1162/jocn.2009.21366

Wang, Z., Xie, L., Lu, T. (2016). Research progress of artificial psychology and artificial emotion in China. *CAAI Transactions on Intelligence Technology, 1*(4), 355-365.

Ward, A. F., Olsen, A. S., & Wegner, D. M. (2013). The harm-made mind: observing victimization augments attribution of minds to vegetative patients, robots, and the dead. *Psychological Science, 24*(8), 1437-1445.

Waytz, A., Cacioppo, J., & Epley, N. (2010).Who sees Human? The stability and importance of individual differences in anthropomorhpism. *Perspectives on Psychological Science, 5*(3), 219-232.

Waytz, A., Morewedge, C.K., Epley, N., Moneleone,G., Jia-Hong, G., & Cacioppo, J. (2010). Making sense by making sentient: Effectance motivation increases anthropomorphism. *Journal of Personality and Social Psychology, 99*(3), 410–35.

Wykowska, A., Chellali, R., Al-Amin, M.M., & Müller, H.J. (2014). Implications of robot actions for human perception. How do we represent actions of the observed robots? *International Journal of Social Robotics*, 6(3), 357-366.

Wynne, C.D.L. (2007). What are animals? Why anthropomorphism is still not a scientific approach to behavior. *Comparative Cognition & Behavior,2*, 125-135. doi:10.3819/ccbr.2008.20008

Yamazaki, R., Nishio, S., Ishiguro, H., Norskov, M., Ishiguro, N., & Balisteri, G. (2012). Social acceptance of a teleoperated android with an embodied communication medium in Denmark. In: S.S. Ge, O. Khatib, JJ. Cabibihan, R. Simmons, MA. Williams (Eds.), *Social Robotics*. ICSR 2012. Lecture Notes in Computer Science (vol 7621). Berlin, Heidelberg: Springer Verlag.

Ziemke, T.& Lowe, R. (2009). On the role of emotion in embodied cognitive architectures: From organisms to robots. *Cognitive Computing, 1*(1), 104-117.

Zlotowski, J., Strasser, E., & Bartneck, C. (2014). Dimensions of anthropomorphism - From humanness to humanlikeness. In *Proceedings of the ACM / IEEE International Conference on Human-Robot Interaction* (HRI '14). New York: USA, (pp.66-73). doi:10.1145/2559636.2559679

7 Appendices

Appendix A: List of Tables

Appendix B: List of Figures

Appendix C: Reception text

Liebe TeilnehmerInnen!

Danke für Ihre Bereitschaft zur Teilnahme an dieser Studie, die Rahmen meiner Masterarbeit an der Fakultät für Psychologie (Universität Wien) durchgeführt wird. Der zeitliche Rahmen für die heutige Befragung wird etwa dreißig Minuten in Anspruch nehmen.

Die Studie behandelt die visuelle Verarbeitung von robotischen Händen. Sie sehen mehrere Videoclips, in denen jeweils ein Mensch eine Manipulation an einem Roboter vornimmt. Bitte bewerten Sie nach jedem Video die drei Fragen, spontan und ohne viel zu überlegen.

Ihre anonym erhobenen Daten werden streng vertraulich behandelt und dienen ausschließlich wissenschaftlichen Zwecken, werden nicht an Dritte weitergegeben und nicht kommerziell genutzt.

Einverständniserklärung:
Durch das Klicken auf "Weiter" bestätigen Sie, dass Sie das vorliegende Informationsblatt gelesen und verstanden haben. Sie erklären sich mit der Teilnahme an dieser Studie sowie mit der Analyse Ihrer Daten durch befugte Personen einverstanden.

Vielen Dank für Ihr Interesse und Ihre Bereitschaft zur Mitarbeit!

Appendix D: Abstract

D.1. English Version

With robots increasingly finding their way to private and work spheres of human lives, it is important to understand under which premises people socially interact with robots. Despite the exponential development in affective computing and the practice of equipping sociable robots with the potential of emotional expression, little is known about the ability and willingness of humans to empathize with robotic agents. The present study explores human empathy towards robots as a function of human-likeness in the physical appearance of robots. By presenting video clips of hostile and friendly interactions of a human hand with robotic hands along an anthropomorphic spectrum, self-reported measures of cognitive and affective empathy were evaluated. A two-way repeated measures ANOVA analysis revealed that people tend to empathize significantly more with robotic entities of a higher anthropomorphic appearance, compared to those with a stronger mechanical look. The hypotheses of significantly higher empathy ratings with increasing human-likeness of technical agents was found for both self-directed (affective) and other-directed (cognitive) aspects of empathy. Contrary to expectations, an influence of differences in personality traits and sociodemographic characteristics on empathy towards robots could not be confirmed.

Keywords: robots, HRI, empathy, anthropomorphism

D.2. German Version

Mit dem zunehmenden Einsatz von Robotern in privaten und professionellen Lebensbereichen stellt sich aktuell die Frage unter welchen Umständen soziale Interaktionen von Robotern und Menschen stattfinden können. Trotz der rapide fortschreitenden Entwicklung künstlicher Emotion und zunehmenden emotionalen Ausdrucksmöglichkeiten sozialer Roboter sind empirische Befunde über die menschliche Bereitschaft und Fähigkeit zu emotionalen Reaktionen auf Roboter rar. Diese Studie untersucht Empathie für Roboter in Abhängigkeit von physischen Merkmalen menschlicher Erscheinungsform. Den

ProbandInnen wurden 32 kurze Videos präsentiert, die soziale und physische Interaktionen einer menschlichen Hand mit robotischen Händen unterschiedlicher anthropomorpher Ausprägung zeigten. Die Hypothese der Studie bezüglich stärkerer empathischer Reaktionen mit zunehmender Menschenähnlichkeit der Roboterhände konnte bestätigt werden. Die durch Selbstbericht erhobenen Empathiewerte der ProbandInnen waren sowohl auf kognitiver als auch auf affektiver Ebene signifikant höher für androide Roboter als für solche mit mechanischerem Aussehen. Die Hypothese eines Zusammenhangs von persönlichen Eigenschaften, soziodemographischen Unterschieden und empathischen Reaktionen konnte in dieser Studie nicht bestätigt werden.

Schlagwörter: Roboter, HRI, Empathie, Anthropomorphismus