Theory of Mind (ToM) on Robots: A Functional Neuroimaging Study

Frank Hegel[1]*, Sören Krach[2,3]*, Tilo Kircher[2], Britta Wrede[1], Gerhard Sagerer[1]

[1] Faculty of Technology, AGAI
Bielefeld University
Universitätsstr. 25
33615 Bielefeld, Germany
fhegel@techfak.uni-bielefeld.de

[2] Department of Psychiatry
RWTH Aachen University Hospital
Pauwelsstr. 30
52074 Aachen, Germany
skrach@ukaachen.de

[3] Central Service Facility
Functional Imaging at the ICCR-Biomat
RWTH Aachen University Hospital
Pauwelsstr. 30
52074 Aachen, Germany

ABSTRACT

Theory of Mind (ToM) is not only a key capability for cognitive development but also for successful social interaction. In order for a robot to interact successfully with a human both interaction partners need to have an adequate representation of the other's actions. In this paper we address the question of how a robot's actions are perceived and represented in a human subject interacting with the robot and how this perception is influenced by the appearance of the robot. We present the preliminary results of an fMRI-study in which participants had to play a version of the classical Prisoners' Dilemma Game (PDG) against four opponents: a human partner (HP), an anthropomorphic robot (AR), a functional robot (FR), and a computer (CP). The PDG scenario enables to implicitly measure mentalizing or Theory of Mind (ToM) abilities, a technique commonly applied in functional imaging. As the responses of each game partner were randomized unknowingly to the participants, the attribution of intention or will to an opponent (i.e. HP, AR, FR or CP) was based purely on differences in the perception of shape and embodiment.

The present study is the first to apply functional neuroimaging methods to study human-robot interaction on a higher cognitive level such as ToM. We hypothesize that the degree of anthropomorphism and embodiment of the game partner will modulate cortical activity in previously detected ToM networks as the medial prefrontal lobe and anterior cingulate cortex.

Categories and Subject Descriptors

J.4 [Social and Behavioral Sciences]: Psychology

General Terms

Design, Experimentation, Human Factors

Keywords

Social Robots, fMRI, Theory of Mind, Anthropomorphism

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
HRI'08, March 12–15, 2008, Amsterdam, Netherlands.
Copyright 2008 ACM 978-1-60558-017-3/08/03...$5.00.

1. INTRODUCTION

As machines become fixtures in the home and workplace, our interactions with them will become more sophisticated and inevitable. Within this context it has been proposed that social robots serve as an interface between humans and technology [1] with the supposition that the more anthropomorphic a robot looks like the more the user will expect the robot to behave like a human being. We assume that a human-like behaving robot is the easiest to use interface simply because humans are highly skilled in having natural interaction with and communication to other humans. Furthermore, users would not have to learn a new technical vocabulary in order to reach a goal when interacting with technical devices. However, these assumptions have barely been tested in a systematic way. Is it indeed the case that the more anthropomorphic a robot looks, the more we expect it to behave in a human-like way? How is this expectation manifested in the human cognitive system and what kinds of expectations are affected? Can we provide evidence for the Uncanny Valley hypothesis? The answers to these questions will have a severe impact on the further development

Fig. 1. Setting of the briefing; from left to right: anthropomorphic robot (AR), computer partner (CP), human partner (HP) and functional robot (FR).

* corresponding authors

of robots as they address the fundamental cognitive mechanisms underlying the interaction with robots. To address these questions we directly investigated the interaction of human participants with artificial/robotic systems with increasing degrees of embodiment and anthropomorphism. Furthermore, participants were scanned by means of functional magnetic resonance imaging enabling us to measure cortical activation during these interactions. We expect the results of our studies to have a severe impact on the design of social robots.

In the experiment participants had to play a version of the classical Prisoners' Dilemma Game (PDG), a paradigm commonly used in social psychology to study aspects of interpersonal behaviour. PDG matrices are used in functional imaging scenarios, because they enable to investigate implicit perspective taking without having confounding influences due to social desirability. First (see Figure 1), the subjects were briefed to play in a face-to-face scenario against a human player (HP), the anthropomorphic robot BARTHOC Jr. (AR), a functional robot (FR) designed with two Lego Mindstorm sets, and a computer laptop player (CP). Afterwards the participants passed on to the MR-scanner located in a neighbouring room and were instructed to play the game during the next 30 minutes. While playing the scanner continuously recorded functional images of the brain.

The responses of the interactors (HP, AR, FR and CP) were randomized in advance, thus ensuring that the subjects' differences in reactions were purely based on the different expectations and perceptions of their interactors. According to the general assumption of social robotics our hypothesis was that the human opponent will evoke strongest activation of brain regions commonly associated with ToM, followed by the anthropomorphic robot, the functionally designed robot and finally the computer (see Figure 1).

In Section II we refer to the interpretation of form and functional images within the field of robotics on the one hand and the concept of ToM on the other hand. Section III describes the design of our fMRI-experiment. The results are presented in Section IV. In conclusion we discuss the paper in Section V.

2. RELATED WORK

In this section we address related work on the form of robots with emphasis on anthropomorphism, the Uncanny Valley hypothesis, and embodiment (Section 2.1). Section 2.2 outlines the related work with regards to the interpretation of Theory of Mind (ToM) and functional images in reference to the field of robotics.

2.1 Form of Robots

According to semiotic theories of product design the form of an object conveys many information [2, 3] about its functionalities or formal aesthetic structures. Most interestingly the form does not only implicitly (I) indicate the use of the object but also provides (II) connotating symbolic information referring to all associations in relation to the object.

(I) Indicating signs of the object's aesthetic form are indicating how specific parts of the object will behave. This idea is very similar to that of affordances [4] but also extends to social interaction. In a similar way as a round tire is indicating the quality of rolling a smile on a face is indicating happiness. This matches with the idea of anthropomorphism. Anthropomorphism is described as a tendency to attribute human characteristics to objects and animals in order to interpret their actions, i.e. functions in an understandable way. According to v. Foerster [5] we anthropomorphize because it allows us to explain things we do not understand in terms that we do understand, and what we understand best is ourselves. Consequently, Duffy [6] argues a robot has to have a certain degree of anthropomorphic attributes for meaningful social interaction.

The form, i.e. the appearance of robots has a substantial influence on the assumptions people have about specific applications and behaviours [7, 8]. For example Fong et al. [9] differentiate between four categories of the robot's form: anthropomorphic, zoomorphic, caricatured, and functional. According to anthropomorphism a human-like robot shape represents a human-like behaviour whereas a functional robot represents the ability to carry out rather restricted tasks. Therefore, the designer of robots should guarantee that the form of a robot matches its functions. In this context DiSalvo et al. [10] suggest to consider a) an amount of robot-ness to emphasise the robot-machine capabilities and to avoid false expectations, b) an amount of human-ness such that the subjects feel comfortable, and c) a certain amount of product-ness such that the robot is also seen as an appliance.

The design of a robot's head is an important issue within human-robot interaction (HRI) because it has been shown that the most non-verbal cues are mediated through the face [11]. Without a face the robot is anonymous [12]. The physiognomy of a robot changes the perception of its human-likeness, knowledge, and sociability. Therefore, people avoid negatively behaving or looking robots and prefer to interact with positive robots [13]. Furthermore, an expressive face indicating attention [14] and imitating the face of a user [15] makes a robot more compelling to interact with.

(II) Due to the form humans are also connotating symbols. The Uncanny Valley hypothesis [16] is dealing with such connotations. The idea of the hypothesis follows from Freud's description of the uncanny (a translation of the German word 'unheimlich') [17]: "derives its terror not from something externally alien or unknown but – on the contrary – from something strangely familiar which defeats our efforts to separate ourselves from it".

The Uncanny Valley represents how an object can be perceived as having enough human-like characteristics to evoke a constrained degree of empathy through one's ability to rationalize its actions and appearance. When the movements and the appearance are almost human-like but not entirely, there are too many expectations of the capabilities and the result is a negative reaction from the observer. In the end, the object becomes so human-like that it is effectively treated as a human being where it has reestablished a balance between anticipated and actual function and form to a sufficient degree that works [1].

Furthermore, the embodiment of a robot may have an effect when interacting with a robot. Bartneck [18] found a facilitation effect in his study with the emotional robot eMuu. Participants acquired a higher score in a negotiation game and they put more

effort into the negotiation when they interacted with the embodied robot character instead of the screen character. This may due to the feeling of social presence [19].

2.2 Theory of Mind / fMRI

As appropriate interaction is more and more becoming a key value in our highly social world, functional neuroimaging studies have addressed this issue in recent times. Thereby, studies increasingly focused on the investigation of human brain activity modulation with respect to the inference of intentions, goals and desires of others. In social cognitive neuroscience inferring the thoughts of a partner has been termed mentalizing or having a Theory of Mind (ToM) [20]. Adopting a ToM is enabling us to predict, anticipate and build a model of the thoughts of a partner and, in some cases, might prepare our behavioural response even before the partner has executed his move yet, an ability proofing to be advantageous in everyday situations (e.g. working environment, private life etc.). Research tasks implicitly evoking ToM related brain activation commonly consist of variations of the classical Prisoners' Dilemma Game (PDG) with subjects mostly asked to either play putative human or computer partners. Recent findings indicate that humans do attribute self-generated actions, intentions and desires rather to human than to computer partners. However, activity in ToM associated brain regions (i.e. medial prefrontal cortex extending into the anterior (para-)cingulate cortex) were reported for the taking over of the perspective of an artificial (computer) partner, too [21, 22, 23, 24, 25, 26]. By now the only functional neuroimaging study incorporating robotic agents stems from Gazzola and colleagues [27]. However, their study addressed the question of whether humans mimic artificial (robotic) limb movements similar to limb movements observed in other humans, thereby focusing on mirror neuron activity with respect to action perception and imitation.

Interestingly, until now it has never been examined whether humans attribute robotic agents higher cognitive processes such as intention or volition. And, if so, whether mentalizing on robotic agents differs at all to mentalizing processes discussed with respect to computer or even to human partners? Especially with respect to a growing interest and demand in appointing robotic agents as domestic help, caretaker or service agents, the question of how we perceive and interact with such artificial agents will be of major concern in future [26, 27, 28, 15, 8, 29, 30].

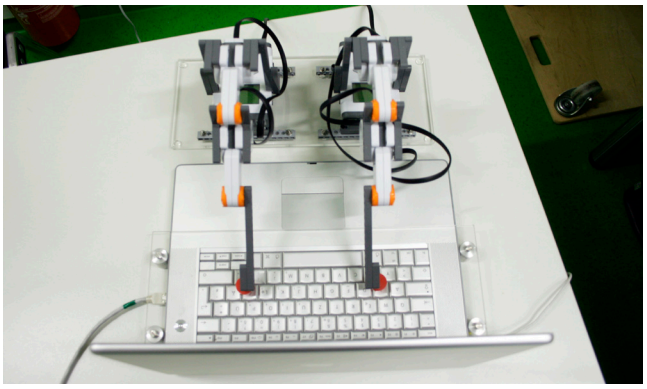

Fig. 2. Functional Robot, Lego Mindstorms

3. EXPERIMENTAL PROCEDURES

3.1 Participants

We present preliminary data of four subjects that participated in the present study (based on the results of these data sets we aim at investigating another 20 participants).

All participants had normal or corrected-to-normal vision and were right-handed according to the Edinburgh Handedness Index [33]. Participants were excluded if they were diagnosed with a past or present psychiatric, neurological, or medical disease. Participants further underwent neuropsychological testing, including attention [34], executive functions [35] and IQ [36]. Furthermore, personality traits were investigated by means of the BFI [37]. The study was approved by the local ethical committee. All participants signed written informed consent prior to participation and were paid a fee for participation.

3.2 Setting of Briefing

Prior to scanning all participants completed a briefing consisting of a "get-together" with their putative game partners: a computer (CP), a functional robot (FR), an anthropomorphic robot (AR) and a human confederate (HP) (see Figure 1).

a) The Functional Robot (FR)

The functional robot with its two arms (see Figure 2) was constructed from two Lego Mindstorm sets (http://mindstorms.lego.com). Each arm consists of two servo motors and a Lego NXT controller that is a computer controlled Lego brick. The two servo motors are directly connected to the NXT controller. The movements of the servo motors are very precise (+/- one degree) so that a believable animation on a computer keyboard is warranted. The behaviour of pressing two buttons on a laptop keyboard is programmed with the Mindstorms NXT software which serves as an intuitive drag and drop programming software to design robots. The functional design represents two arms modeled after a human arm to support the idea of anthropomorphism.

b) The Anthropomorphic Robot BARTHOC Jr.

BARTHOC Jr. looks like a child at the age of five years with the size of 65 cm from the waist upwards (see Figure 3). The robot is able to move its torso which is mounted on a 65 cm high chair-like socket to the left and to the right. The socket includes the power supply, actuator controllors so-called iModules, and two serial interfaces to a computer. One interface controls the head and neck actuators, the other one is connected to the actuators below the neck.

In total 41 actuators consisting of DC- and servo motors move the robot. The face has ten degrees of freedom to control jaw, mouth angles, eyes, eye brows, and eye lids. Therewith, the robot is able to imitate human-like facial expressions. The eyes are

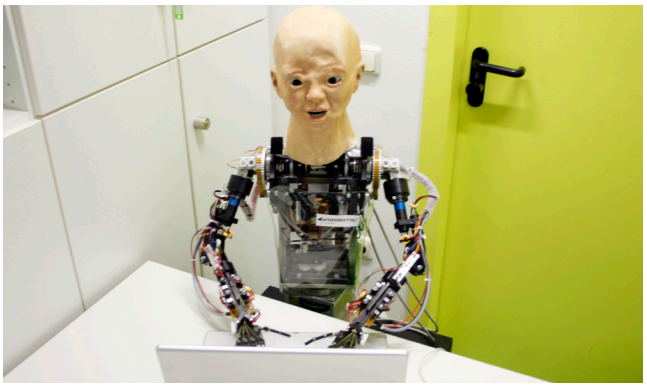

Fig. 3. Anthropomorphic Robot, BARTHOC Jr.

vertically aligned and horizontally moveable. Each eye contains a Firewire color video camera with a resolution of 640 x 480 pixels. Furthermore, the head can be turned, tilted to its side and slightly shifted forwards and backwards. Each arm can be moved similarly to the movements of a human arm. With its five fingers on each hand BARTHOC Jr. is able to show simple grips as well as deictic gestures. The fingers use only one bending actuator, but they are autonomously controllable for believable movements. By using different facial masks which are made of latex we are able to alter the appearance of BARTHOC Jr's face. This enables us to use male and female personalities within specific settings to study gender effects. For extended experiments with an adult-like robot we use the taller robot BARTHOC [29].

After introducing participants to their opponents for the upcoming game sessions, participants were seated face-to-face with their anticipated game partners (see Figure 1). A notebook placed in front of the participant displayed the instruction of the experiment. Notebooks – placed in front of each interactor - were linked to the participant's notebook via mocked connecting cables. The keyboards placed in front of the robots were covered by a custom-made plexiglass device. This construction was equipped with a fixed two-button system allowing the robot to press only two single keys (see Figure 2).

For the briefing both robots were programmed in advance to push their keyboard buttons exactly at the same time when the subjects believed to play them. Similarly, the confederate contemporaneously pressed his buttons when the subject assumed playing the human partner. However, during the tutorial as well as during the entire experiment, the response behaviour of the subject's partners was randomized, thereby not enabling participants to really cooperate or find "a best way". By this means it will be possible to infer pure "intentional stance" associated neural activity [38, 39] as possible strategies of the putative partners are hold constant.

3.3 Stimuli and Task Design

The briefing comprised two tutorial rounds for each condition (CP, FR, AR, HP and baseline). The task resembled decision games already applied by other research groups and can be considered as a variant of the PDG [25, 26, 22, 23, 24].

Taken together, participants always had to find a decision about cooperating or defecting with the respective interactor in a series of nine games in a row. Cooperation was signalled by pressing the left button on the computer keyboard, defection by pressing the right button, respectively. Depending on the interactor's decision, the participant immediately received a previously defined and explicitly learned pay-off feedback, making the scenario highly interactive.

The payoff feedback, as conveyed by the decision matrix (specifically developed and pre-tested) worked as follows: If both contenders were pressing the left button, both of them would be gratified with 20 points each (CC). In case that the participant would press the left button (cooperate) with the respective partner pressing the right button at the same time (defect), the participant would return with 10 points for this game, but the opponent would receive 20 points (CD). The other way around, the participant (defect) would reap 20 points, while the opponent would receive 10 points (DC). In case both contenders choose to defect, the dilemma would eventuate with both sides receiving zero points (DD). CC implies mutual cooperation, while DD involves mutual non-cooperation [26]. Games were interspersed by a low-level baseline condition enforcing participants to alternately press the right and left button when a central cross hair appeared on the computer screen.

Importantly, the instruction given to the participants involved the demand to both, "win a series of games and reach a virtual highscore". As these two converse goals could, per definition, not be reached by solely pressing one button, this matrix secured an almost equal pressing of both buttons, thereby supporting the idea behind: to find a decision based upon the reasoning about the opponent's last decisions, i.e. triggering Theory of Mind (ToM).

Finally, the briefing pursued two goals: firstly, familiarizing participants with the decision matrix and secondly, triggering a strong attachment of the participants to their game partners.

3.4 fMRI Setting

After the briefing the participant passed on to the MR-scanner located in an adjacent room. The experimenter gave last instructions and clarified that the participant understood the winning matrix as well as the demand to both "win a series and reach a virtual highscore". All putative game partners remained seated, while their notebooks were demonstratively "connected" to the MR presentation computer.

At this point of time the help of the confederate was not needed anymore. With the beginning of the functional imaging recording a randomized script file (the experiment was performed using Presentation® software; Version 0.70, www.neuro-bs.com) was started with the projection of the identical setting of the briefing onto the MR-compatible video goggles (Resonance Technology). Prior to each series of games, participants were informed about the partner being played in the following (via portraits of CP, FR, AR, HP or baseline; 2500ms; see Figure 4).

Hereafter, a central cross on the screen indicated the start of a series. In order to indicate a decision (either cooperation or defection), participants had to press one of two buttons with their right

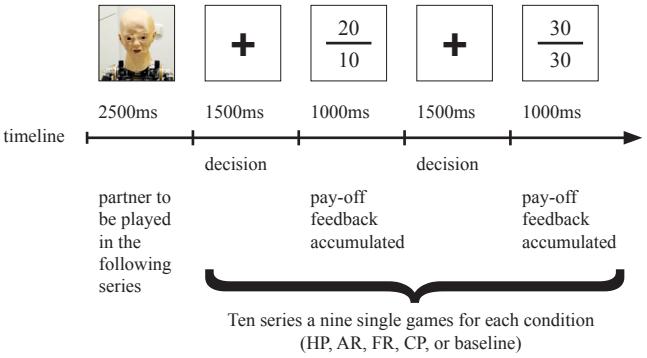

Fig. 4. Stimuli display and time course of the applied paradigm

hand on a fiberoptic custom-made and MR compatible response box. The central cross disappeared after 1500ms and was followed by an accumulated pay-off feedback for the current series (2000ms) enabling participants to draw an exact inference about the response selection of the current partner. The participant's pay-off was indicated by the lower numbers, the partner's pay-off by the upper numbers, respectively (see Figure 4). During the low level baseline no numeral response feedback was given, instead two crosses replaced the numbers on the upper and lower side of the bar.

The outcomes of each single game were recorded and saved to a computer file. A series of nine games per condition completed one block. Overall, participants played ten blocks per condition. After scanning participants were asked to fill out a final questionnaire about their impressions of the task and their opponents.

3.5 Image Acquisition and Analysis

All scans were performed on a 3T whole body scanner (see Figure 5; Phillips Medical Systems, Achieva, Best, Netherlands) using standard gradients and a standard quadrature head coil. Participants lay in a supine position, while head movement was limited by foam padding within the head coil. In order to ensure optimal visual acuity participants were offered MR-compatible glasses that could be fixed to the video glasses. For each participant, we acquired one series of 870 EPI-scans, lasting in total about 28.3 minutes. Stimuli were presented in a blocked design fashion, with ten blocks per condition and a block length of nine single games (one single game lasting 3500ms).

Scans covered the whole brain, including five initial dummy scans parallel to the AC/PC line with the following parameters: number of slices (NS): 32; slice thickness (ST): 3.5 mm; interslice gap (IG): 3.75 mm; matrix size (MS): 64x64; field of view (FOV): 192 mm x 192 mm; repetition time (TR): 2000 ms; echo time (TE): 30 ms; flip angle (FA): 90°.

For anatomical localization, we acquired high resolution images with a T1-weighted 3D FFE sequence (TR = 9.896 ms; TE = 4.6 ms; NS = 180 (sagital); ST = 1 mm; IG = 0 mm; FOV = 256 x 256 mm; voxel size = 1 x 1 x 1 mm).

MR images were analyzed using Statistical Parametric Mapping (SPM5, ww.fil.ion.ucl.ac.uk) implemented in MATLAB 7.0 (Mathworks Inc., Sherborn, MA, USA). After discarding the first five volumes, all images were realigned to the first image to correct for head movement. Unwarping was used to correct for the interaction of susceptibility artefacts and head movement. Volumes were then normalized into standard stereotaxic anatomical MNI-space by using the transformation matrix calculated from the first EPI-scan of each participant and the EPI-template. Afterwards, the normalized data with a resliced voxel size of 4x4x4 mm were

Fig. 5. MR-Scanner within the Experiment

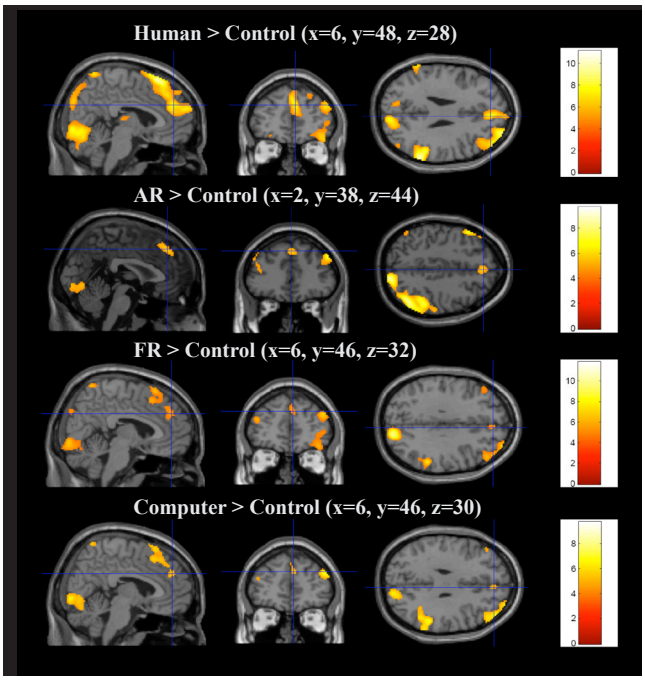

Fig. 6. Subject 3; all four conditions vs. control, respectively. Crosshair located at the local maxima activation. Activated areas comprise the right medial frontal gyrus (BA 6/9) extending onto the superior frontal gyrus (BA 8/9). (threshold of p > .05, FWE)

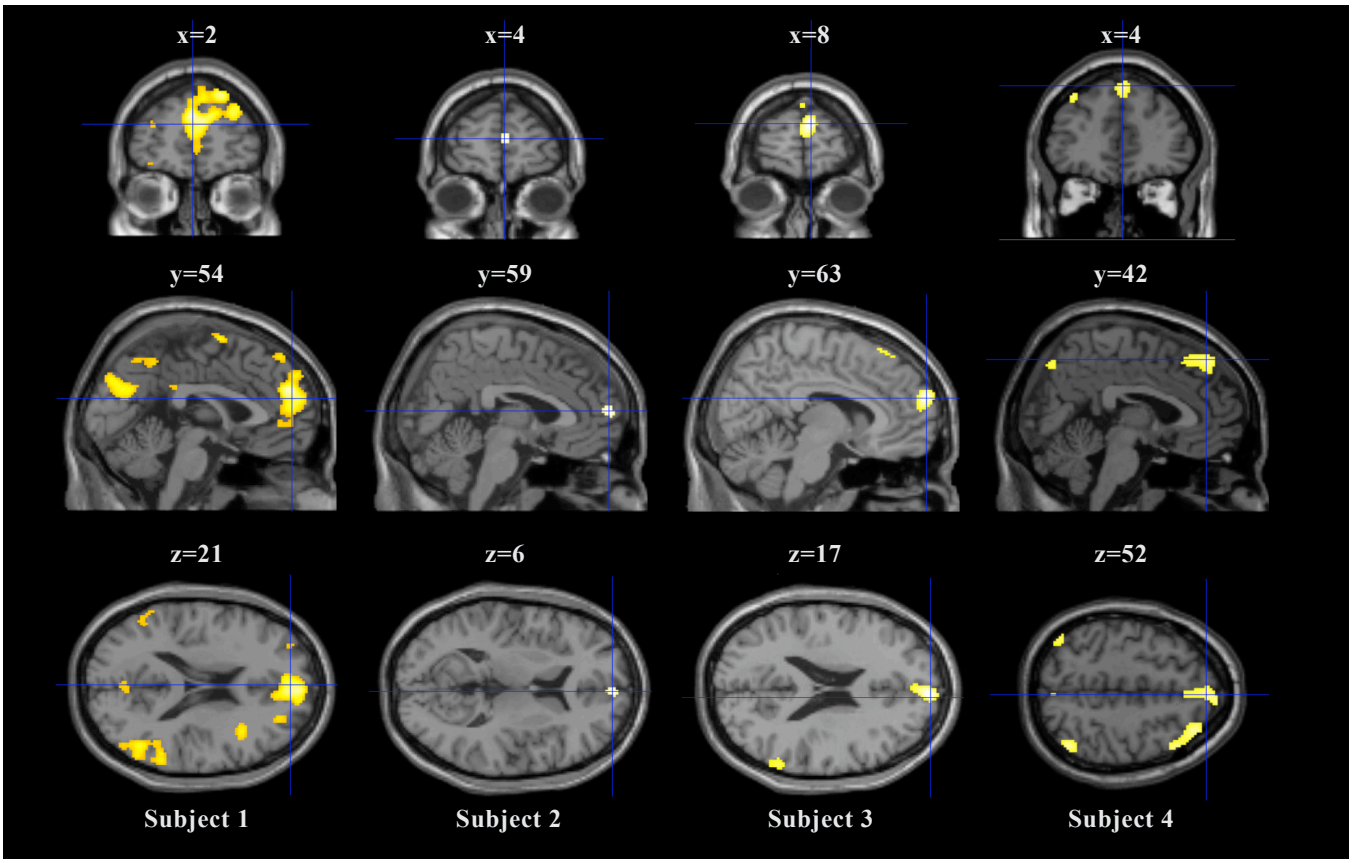

Fig. 7. Complex contrast "human vs. robots" (AR & FR); figures display activity measures of all four subjects. The crosshair is located at the local maxima activation (medial frontal cortex (BA 9/10) extending into the superior frontal cortex). (threshold of p > .05, FWE)

smoothed with an 8-mm FWHM isotropic Gaussian kernel to accommodate inter-participant variation in brain anatomy. The time series data were band-pass filtered to remove artefacts due to cardio-respiratory and other cyclical influences.

A general linear model (GLM) comprising five conditions (CP, FR, AR, HP and baseline) was specified for each participant. For all analyses we applied a conservative voxel-wise threshold of $p<0.05$ (FWE equaling Bonferoni correction). The reported voxel coordinates of activation peaks were transformed from MNI space to Talairach & Tournoux atlas space [40] by non-linear transformations (www.mrc- cbu.cam.ac.uk/Imaging/mnispace.html).

4. RESULTS / NEUROIMAGING DATA

The neuroimaging results will only be discussed with respect to frontal activity modulation (for more details on the functional neuroimaging data of this study the reader is referred to Sören Krach [skrach@ukaachen.de]).

Regarding highly preliminary single subject data (see Figure 6; exemplarily subject 3) it can be stated that all four experimental conditions (compared to the low-level baseline) elicited similar medial prefrontal and superior frontal cortex activations (BA 8/9/10) (FWE corrected at $p>.05$). As these data sets must be regarded as highly preliminary (N=4) no prediction about population effects can be derived so far. However, all subjects exhibit profound activity in brain regions considered to be participating in common Theory of Mind (ToM) tasks.

Regarding neural differences with respect to playing a real human partner relative to playing robots (FR & AR), a stable and persisting medial frontal cortex activation extending into the superior frontal cortex manifests for each single subject (FWE corrected at $p>.05$). The local maxima activations are nearly identical to the centre of activation exhibited during the simple contrasts and correspond perfectly to the brain regions generally associated with ToM (see Figure 7).

5. DISCUSSION

From a methodological point of view, misleading the participants by displaying random responses that are attributed to different human and computer opponents enables to calculate hemodynamic changes related to differences in the instruction only, ruling out possible interaction effects of scattered strategic alliances. Hence, the present paradigm offered the possibility to uniquely measure

brain activity related to the simple supposition made by the participants about the intentions, goals and ambitions of the partner independent of its behavioural response [39].

The design of the experimental setting exposes to be highly believable for future studies. None of the four subjects noticed that the behaviour of all interactors was purely random at any time.

On a neuronal level, we could demonstrate that participants tried to figure out the goals and intentions of all interactors, documented by highly significant activations of brain regions commonly associated with mentalizing, i.e. the medial prefrontal cortex extending into the anterior cingulate cortex (ACC) [26, 24, 23, 39, 22] (e.g. Figure 5).

Furthermore, all four subjects displayed strong cortical activity in these "suspect" areas when encountered with the human opponent relative to the robotic interactors (see Figure 7). However, given the limited number of subjects (N=4) it was not possible to calculate more detailed contrasts, as e.g. robotic agents > computer, FR > AR or vice versa. Yet, careful but rather preliminary inspection of the data revealed that three out of four subjects empathized more with the robots (AR or FR) than with the laptop indicating that with a larger sample size we might find a significant effect of anthropomorphism on the activation of ToM-related areas. We did not find a consistent tendency of the anthropomorphic robot (AR) evoking stronger reactions than the functional robot (FR) speculatively indicating that the anthropomorphic robot might cause an Uncanny Valley effect. To predict distinctive statements in reference to anthropomorphic effects, the Uncanny Valley hypothesis, and effects of embodiment we are continuing our studies by expanding the highly expensive experiment with more than 20 subjects to consolidate our findings and hypotheses.

In summary, although no statistically significant results with respect to our hypothesis of anthropomorphism causing higher ToM related activity could be found, we were able to show that our setting was stable and provided high face validity. With an increasing sample size, we will be able to test each of our hypotheses. Secondly, we could show sustainable activities within the Theory of Mind (ToM) network while participants were interacting with each partner: a human, an anthropomorphic robot, a functional Lego Mindstorms robot, and a computer laptop. We were thus able to show a tendency towards higher activity modulation when participants were facing robotic partners relative to the laptop partner. If this trend can be confirmed with increasing sample size, it would implicate that by designing robots with embodied/anthropomorphic features human interaction partners would indeed expect human-like behaviour. This is consistent with the idea that physically embodied agents will evoke stronger emotional responses by the users because they are able to physically manipulate the environment thus indicating that physical movement as well as the presentation of internal reasoning processes need to be very carefully designed in order to be acceptable and comfortable for human users. Finally, we could demonstrate a highly robust and significant "human superiority effect".

With the present study we provide a new methodology to analyze the basic mechanisms of mentalizing and a method to investigate the basis of acceptance and comfort factors caused by the design of robots' appearances and behaviors.

6. ACKNOWLEDGEMENTS

We would like to address thanks to Sebastian Gieselmann, Stephan Buschkämper, Sabrina Oehlschläger, and Alessandra Mainieri for their support realizing this study.

This work has been supported by the EU funded IP-project COGNIRON and the German Research Foundation (DFG) within the SFB 673 'Alignment in Communication' – Project C2 (Communicating Emotions).

7. REFERENCES

[1] B. Duffy and G. Joue (2004). 'I, robot being', Intelligent Autonomous Systems Conference (IAS8), Amsterdam, The Netherlands.

[2] D. Steffen (1997). 'On a Theory of Product Language, Perspectives on the hermeneutic interpretation of design objects', in: formdiskurs, Journal of Design and DesignTheory, Nr. 3, 2/ 1997

[3] D. Steffen (2000). 'Design als Produktsprache, Der Offenbacher Ansatz in Theorie und Praxis', Frankfurt am Main.

[4] J. J. Gibson (1977). 'The Theory of Affordances'. In Perceiving, Acting, and Knowing, Eds. Robert Shaw and John Bransford.

[5] H. v. Foerster (1997). 'Wissen und Gewissen. Versuch einer Brücke'. Frankfurt.

[6] B. Duffy (2003). 'Anthropomorphism and The Social Robot', Special Issue on Socially Interactive Robots, Robotics and Autonomous Systems 42 (3--4).

[7] J. Goetz, S. Kiesler, and A. Powers (2003) 'Matching robot appearance and behavior to tasks to improve human-robot cooperation', in Proceedings of the 12th IEEE Workshop on Robot and Human Interactive Communication (RO-MAN 03), pp. 55–60, San Francisco, CA.

[8] M. Lohse, F. Hegel, A. Swadzba, K. J. Rohlfing, S. Wachsmuth, and B. Wrede (2007). 'What can I do for you? Appearance and Application of Robots'. AISB Workshop on The Reign of Catz and Dogz? The role of virtual creatures in a computerised society, Newcastle, GB.

[9] T. W. Fong, I. Nourbakhsh, and K. Dautenhahn (2002). 'A Survey of Socially Interactive Robots: Concepts, Design, and Applications', Robotics and Autonomous Systems, 42(3 – 4), 142 – 166.

[10] C. F. DiSalvo, F. Gemperle, J. Forlizzi, and S. Kiesler (2002). 'All robots are not created equal: the design and perception of humanoid robot heads', in DIS '02: Proceedings of the conference on Designing interactive systems, pp. 321–326, New York, NY, USA. ACM Press.

[11] M. Blow, K. Dautenhahn, A. Appleby, C. L. Nehaniv, and D. C. Lee (2006). 'Perception of Robot Smiles and Dimensions for Human-Robot Interaction Design' 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 06), pages 469-474, Hatfield, UK.

[12] J. Donath (2001). 'Mediated Faces'. In: M. Beynon, C. L.Nehaniv, K. Dautenhahn (Eds.). Cognitive Technology: Instruments of Mind. Proceedings of the 4th International Conference, CI 2001, Warwick, UK.

[13] R. Gockley, J. Forlizzi, and R. Simmons (2006). 'Interactions with a moody robot', in HRI'06: Proceeding of the 1st ACM SIGCHI/SIGART conference on human-robot interaction, pp. 186–193, New York, USA. ACM Press.

[14] A. Bruce, I. Nourbakhsh, and R. Simmons (2001). 'The role of expressiveness and attention in humanrobot interaction', in Proc. AAAI Fall Symp. Emotional and Intel. II: The Tangled Knot of Soc. Cognition.

[15] F. Hegel, T. Spexard, T. Vogt, G. Horstmann, and B. Wrede (2006). 'Playing a different imitation game: Interaction with an Empathic Android Robot', in Proc. 2006 IEEE-RAS International Conference on Humanoid Robots (Humanoids 06), pp. 56–61. IEEE.

[16] M. Mori (1997). 'The Buddha in the Robot'. Charles E. Tuttle Co.

[17] S. Freud (1953). 'The Uncanny'. The Standard Edition of the Complete Psychological Works of Sigmund Freud, ed. & trs. James Strachey, vol. XVII (London: Hogarth), pp. 219-252.

[18] C. Bartneck (2003), 'Interacting with and embodied emotional Character'. Proceedings of the DPPI2003 Conference, Pittsburgh, pp. 55-60.

[19] Y. Jung and K. M. Lee (2004). 'Effects of physical embodiment on social presence of social robots'. Proceedings of Presence 2004, 80-87.

[20] D. Premack and G. Woodruff (1978). 'Does the chimpanzee have a theory of mind?'. Behavioural and Brain Science, 1, 515-526.

[21] J. Decety, P. L. Jackson, J. A. Sommerville, T. Chaminade, and A. N. Meltzoff (2004). 'The neural bases of cooperation and competition: an fMRI investigation'. Neuroimage, 23(2), 744-751.

[22] H. Fukui, T. Murai, J. Shinozaki, T. Aso, H. Fukuyama, T. Hayashi, et al. (2006). 'The neural basis of social tactics: An fMRI study'. Neuroimage, 32(2), 913-920.

[23] T. Kircher, T. Lataster, D. Majoram, I. Blümel, L. Krabbendam, J. Weber, et al. (in review). Online measurement of Theory of Mind (ToM).

[24] S. Krach, I. Blümel, D. Majoram, T. Lataster, L. Krabbendam, J. Weber, et al. (in review). 'Playing the game differently - sex differences in mentalizing detected with functional neuroimaging'.

[25] J. K. Rilling, D. Gutman, T. Zeh, G. Pagnoni, G. Berns, and C. Kilts (2002). 'A neural basis for social cooperation. Neuron'. 35(2), 395-405.

[26] J. K. Rilling, A. G. Sanfey, J. A. Aronson, L. E. Nystrom, and J. D. Cohen (2004). 'The neural correlates of theory of mind within interpersonal interactions'. Neuroimage, 22(4), 1694-1703.

[27] V. Gazzola, G. Rizzolatti, B. Wicker, and C. Keysers (2007). 'The anthropomorphic brain: the mirror neuron system responds to human and robotic actions'. Neuroimage, 35(4), 1674-1684.

[28] J. Fritsch, B. Wrede, and G. Sagerer (2005). 'Bringing it all together: Integration to study embodied interaction with a robot companion'. Paper presented at the AISB 2005 Symposium - Robot Companions: Hard Problems and Open Challenges in Robot-Human Interaction, Hatfield, England.

[29] M. Hackel, M. Schwope, J. Fritsch, B. Wrede, and G. Sagerer (2006). 'Designing a sociable humanoid robot for interdisciplinary research'. Advanced Robotics, 20(11), 1219-1235.

[30] M. Hackel, S. Schwope, J. Fritsch, B. Wrede, and G. Sagerer (2005). 'A humanoid robot platform suitable for studying embodied interaction'. Paper presented at the RSJ Int. Conf. on Intelligent Robots and Systems, Edmonton, Alberta, Canada.

[31] T. Spexard, A. Haasch, J. Fritsch, & G. Sagerer (2006). 'Human-like person tracking with an anthropomorphic robot'. Paper presented at the IEEE Int. Conf. on Robotics and Automation (ICRA), Orlando, Florida.

[32] B. Wrede, M. Kleinehagenbrock, & F. Fritsch (2007). 'Towards an integrated robotic system for interactive learning in a social context'. Paper presented at the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems - IROS 2006, Bejing.

[33] R. C. Oldfield (1971). 'The assessment and analysis of handedness: the Edinburgh inventory'. Neuropsychologia, 9(1), 97-113.

[34] R. Brickenkamp (2002). 'Der Aufmerksamkeits-Belastungstest d2'. Göttingen: Hogrefe.

[35] R. Reitan, & D. Wolfson (1985). 'The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation'. Tucson: Neuropsychology Press.

[36] S. Lehrl (2007). 'Der Mehrfachwahl- Wortschatz- Intelligenztest'. Göttingen: Hogrefe.

[37] O. P. John (1990). 'The "Big Five" factor taxonomy: Dimensions of personality in the natural language and in questionnaires'. In L. Pervin (Ed.), Handbook of personality: Theory and research (pp. 66-100). New York: Guilford Press.

[38] D. C. Dennett (1987). 'The intentional stance'. Cambridge, MA: The MIT Press.

[39] H. L. Gallagher, A. I. Jack, A. Roepstorff, and C. D. Frith (2002). 'Imaging the intentional stance in a competitive game'. Neuroimage, 16(3 Pt 1), 814-821.

[40] J. Talairach, and P. Tournoux (1988). 'Co-planar stereotaxic atlas of the human brain'. Stuttgart, Germany: Thieme.