Note: this is the accepted version of the manuscript. The final publication is available at: http://link.springer.com/article/10.1007/s12369-014-0239-x

Implications of robot actions for human perception. How do we represent actions of the observed robots?

Agnieszka Wykowska[1,2] · Ryad Chellali[3] · Md. Mamun Al-Amin[1,4] · Hermann J. Müller[1,5]
[1] General and Experimental Psychology Unit, Dept. of Psychology, Ludwig-Maximilians-Universität, Leopoldstr. 13, 80802 Munich, Germany
[2] Institute for Cognitive Systems, Technische Universität München, Karlstr. 45/II, 80333 Munich, Germany
[3] Istituto Italiano di Tecnologia-PAVIS, Via Morego, 30, 16165 Genova, Italy
[4] Department of Pharmaceutical Sciences, North South University, Plot-15, Block-B, Bashundhara, Dhaka, 1229, Bangladesh
[5]Birkbeck College, University of London, Malet Street, Bloomsbury, London WC1E 7HX UK

{agnieszka.wykowska@psy.lmu.de, ryad.chellali@iit.it, bd_pharmacy@yahoo.com, hmueller@psy.lmu.de}

Abstract Social robotics aims at developing robots that are to assist humans in their daily lives. To achieve this aim, robots must act in a comprehensible and intuitive manner for humans. That is, humans should be able to cognitively represent robot actions easily, in terms of action goals and means to achieve them. This yields a question of how actions are represented in general. Based on ideomotor theories [1] and accounts postulating common code between action and perception [2] as well as empirical evidence [3], we argue that action and perception domains are tightly linked in the human brain. The aim of the present study was to examine if robot actions would be represented similarly, and in consequence, elicit similar perceptual effects, as representing human actions. Our results showed that indeed robot actions elicited perceptual effects of the same kind as human actions, arguing in favor of that humans are capable of representing robot actions in a similar manner as human actions. Future research will aim at examining how much these representations depend on physical properties of the robot actor and its behavior.

Keywords *Action-Perception Links · Perceptual Processing · Human-Robot Interaction · Action Understanding*

1 Introduction

The field of social robotics aims at designing artificial agents that will help people in their daily lives and as such, these robots should be part of humans social sphere. Realization of this aim depends on whether humans will consider robots as socially accepted partners or simple machines. If robots are to be perceived as machines or only simple automata, it is enough to consider - for robotic designs - only the functions they support and the ways to access these functions (human-machine interface). However, for a robot to be perceived as a social partner with which natural interactions are possible, mechanisms underlying social cognition in the human mind also need to be taken into account.

One of the key aspects of social cognition is understanding others' actions and their action goals. Humans have developed mechanisms, such as mentalizing/theory of mind [4, 5] or simulating [6, 7] that allow action understanding. Proponents of the mentalizing/theory of mind mechanism argue in favor of a higher-order cognitive process, which underlies understanding/explaining actions with reference to other people's mental states. That is, in order to explain behavior B of an agent A, one refers to the underlying mental states, such as beliefs, desires or intentions: A does B because A desires C. On the other hand, proponents of the so-called simulation theory argue that when observing actions, one automatically simulates similar actions in their own cognitive system, which should also allow for action understanding. While the first class of mechanisms might be considered more explicit and reflective, the second class is presumably more implicit and reflexive [8]. In this paper, we will address the implicit, reflexive, and thus presumably more fundamental mechanisms underlying social cognition – that is, the mechanisms involved in action simulating. In particular, we will focus on the perceptual consequences of representing observed (and then reproduced) actions.

1.1. Simulating when observing: the impact of action on perception and the intentional weighting mechanism.

If observed actions are simulated in observer's own cognitive system (actions are mapped onto one's own action repertoire), then the explanation of how others' actions are represented is highly dependent on understanding of how actions are represented in general. In line with the ideomotor theory [1], accounts postulating common code for action and perception [2], predictive coding framework [9] or forward models of motor control [10], we argue that action and perception domains are tightly linked in the human brain. In previous work, a so-called ***intentional weighting*** mechanism has been

postulated [2-3, 11-14] – a mechanism, which presumably tunes perceptual processing to action representations. The idea [11] is the following: when actions are represented in the human brain with the prospect of potential action production, the representations have two components, (i) offline representation where invariant characteristics of an action are specified and stored in memory (e.g., effector with which the action is to be performed), and (ii) a more flexible representation consisting in open parameters which are specified during online control (e.g. particular orientation of a hand or size of grip aperture). Since several parameters of an action are often open and need to be specified during online action control in real-life situations, perceptual system needs to deliver information to the online control in an efficient manner. We postulate that the intentional weighting mechanism serves this purpose: the purpose tuning perceptual processes to action representations. Through lifelong experience, the human brain learns which perceptual characteristics might become relevant for which actions, and during action planning prioritizes processing of the potentially relevant perceptual features.

To give an example: if one represents an action of catching an object, one can specify the effectors (arms) that need to be used offline; and thus representing this action will elicit representation of activating motor commands to these effectors. However, specific orientation of the hand, or grip aperture cannot be specified offline and needs to be adjusted online during action control. It is precisely for this purpose that intentional weighting mechanism operates: through tuning perception to action-relevant characteristics (i.e., prioritizing processing of orientation, size, etc. while filtering out processing of action-irrelevant characteristics such as color in this case), the brain facilitates delivery of sensory information essenital for efficient online control.

Therefore, observing an action, which elicits a certain representation should - according to this theoretical standpoint - evoke not only motoric representation of the action but should also elicit sensory processes that are tuned to the represented action. These should be the underlying mechanisms of simulating when observing.

In our previous research [3, 12, 13], we examined behavioral manifestations of the intentional weighting mechanism and their neuronal correlates [14]. In a series of studies, participants were asked to perform a perceptual task: a visual search task (detection of a target that differed from the surrounding distractors by only one feature, see Fig. 1) and a movement task (grasping or pointing to an object situated elsewhere than the objects of the visual search task). Importantly, the visual search targets were defined either by size or luminance dimension, thereby creating two action-perception ***congruency*** pairs: size was a congruent dimension with grasping while luminance was a congruent dimension with pointing. The key feature of the paradigm was that the perceptual task was entirely unrelated to the movement task both perceptually and motorically: the perceptual task was concerned with objects (circles) presented on the computer screen, while the movement task was performed on objects placed below the screen. Also responses were unrelated, as the perceptual task was performed by means of pressing a key on a standard computer mouse executed with two fingers of a dominant hand; while responses in the movement task were executed by grasping/pointing with the other hand to actual objects. Results repeatedly showed [3, 12-14] action-perception ***congruency effects***: action-congruent dimensions in the perceptual task were better detected (faster reaction times; better accuracy) than action-incongruent dimensions. Moreover, Event-Related Potentials (ERPs) of the human electroencephalography (EEG) showed modulation of early attention-related ERP components, related to action planning [14]. Since the tasks were unrelated, the congruency effects were most likely due to overlap between action and perception domains at the representational level in the brain. We concluded that the action-related intentional weighting mechanism weighted action-relevant dimensions higher, thereby allowing prioritized processing of those dimensions, and as a result, better performance.

1.2. Aim of the present study

The aim of the present study was to examine if the intentional weighting mechanism, and its behavioral manifestations (the action-perception congruency effects) would also be found when participants observe robot actions. In previous research, the to-be prepared actions were signaled to participants by means of a picture of a human hand performing the respective action. In order to perform the task, participants needed to understand the observed action and represent it in their cognitive system to plan and execute the action properly.

In the present study, we examined whether humanoid robot actions map onto action representations in the human brain and thereby to test in an implicit manner human's ability for understanding robot actions. This study was aimed as the first step to systematically examine how various shapes of a robot and the degree of resemblance to a human influence whether the robot's action can evoke appropriate action representations in the human brain. The robot's shape is assumed to be one of the critical characteristics of an entity (robot) that influences the impressions of a human regarding that entity. The uncanny valley hypothesis [15], however, suggests a limit beyond which resemblance of an artificial machine to a human can become repulsive

due to being almost-like human but still different. The uncanny valley hypothesis has been proposed in the 70's and addressed from different angles. In [16] for instance, the authors proposed the predictive coding framework, which interprets the uncanny valley from the biological perspective: uncanny valley is explained as the presence of discreapencies between actual observations and perceptual expectations. Other authors [17] hypotesize that the uncanny valley is not a unique phenomenon and may be caused by several factors including cultural background. Following this line of argument, [18] suggest that the uncanny valley is specifically a generational phenomena which is perhaps more relevant to older people, and which deos not apply to young generations who are better used to information technologies (computer graphics, video games, etc.). On the other hand, proponents of the *media equation* theory argue that natural and social interaction with media as well as personification of technology is quite a universal phenomenon [19]. In any case, this broad spectrum of various types of hypotheses, theories and interpretations does not allow clear and useful conclusions about Mori's intuition. In our approach, we aim at providing a methodology for testing the implicit and fundamental mechanisms of human social cognition during observation and interaction with others, and in particular, with robots.

To meet this aim, we designed an experimental paradigm similar to the one used in studies reported previously [3, 12-14] with the crucial difference of using cartoon pictures of human hands and humanoid robot hands (instead of human hands), which signaled the required movement (grasping/pointing). We hypothesized that if humanoid robot actions are represented in a similar manner to other human's actions, then we should be able to observe action-perception congruency effects in both the robot and the human conditions. In contrast, observing congruency in only one but not the other condition would indicate that representation of observed action depends on the type of the actor being observed. A similar question has been addressed in [20]. The authors compared visuomotor priming effects for actions signaled by human and robotic (pincer) hands. Crucially, in our approach we intend to extend these findings by examining *perceptual* consequences of action (re-) production, and not the effects at the level of action production itself; and most importantly, we plan to systematically examine the continuum ranging from human-like hand shapes to shapes resembling human hands to a lesser and lesser degree. As the first step, however, we employed robot hands that resembled human hands at large, both in morphology as well as functionality that the morphology implied.

2 Methods

2.1 Participants

Twenty healthy volunteers (7 women; age range: 18-30 years; mean age = 24.25) took part in the experiment. Two participants were left-handed but they were using the computer mouse in the same way as right-handers. The participants were naive with respect to the purposes of this experiment. All participants had normal or corrected to normal vision and have provided informed consent regarding participation in the experiment.

2.2 Experimental design

Participants performed a visual search task for a target defined either by size or by luminance, see Fig. 1. The visual search task meant that participants were to detect a target when it was presented on the display (positive response) and reject target absent trials with a negative response. The luminance target was always lighter (Fig. 1 left) while the size target was always larger than the distracters (Fig. 1 right). In all trials, participants were also asked to perform a movement task (grasping or pointing) according to a picture of a human cartoon hand or a robot hand, see Fig. 2. Importantly, the movement was not supposed to be executed until the completion of the visual search task (see the Procedure section below). Thereby, during processing of the perceptual task (the visual search), participants had the movement representation activated. With such a design, we created two action-perception congruency pairs: size was the congruent dimension with the grasping movement (in grasping, one needs to specify size of object for relevant the grip aperture, see also [3]), while luminance was congruent with the pointing movement (luminance is a good hint for localization, and in pointing, localization is crucial, see [3, 21]). We expected to observe action-perception congruency effects in the visual search task. As the dependent variables were reaction times (RTs) and error rates, we expected that detection of size targets would be better (faster RTs and/or lower error rates) in the grasping condition relative to pointing; while detection of luminance targets would be better in pointing relative to grasping.

2.3 Stimuli and Apparatus

Stimuli were presented on 19 inch CRT screen, with a 100 Hz refresh rate placed at a distance of 100 cm from an observer. Responses were registered with a Logitech optical mouse. The whole experiment was programmed in E-Prime (Psychology Software Tools, Inc.) run on an Intel ®CoreTM 2 CPU 6700 @ 2.66 GHz).

2.3.1. Visual Search: The visual search displays consisted of three imaginary circular arrays of 6.8°, 4.8°and 2.8°diameter, with 16, 8 and 4 visual search items, respectively. The target item was always presented in the middle circle of the array.

Luminance targets: All search items were of the same size (1.1˚). The target item was always lighter (luminance: 58 cd/m^2) than the other circles, see Fig. 1, left. Size targets: Size-target search display comprised of 28 grey circular items (1.1˚of visual angle; 15 cd/m^2 of luminance). The target item was always a larger circle (1.4˚of diameter), see Fig. 1, right. All displays were shown on a light-gray background, and in the target-absent trials, all items were identical (1.1˚of visual angle; 15 cd/m^2 of luminance). There were 50% of target present and 50% of target absent trials in both luminance and size blocks.

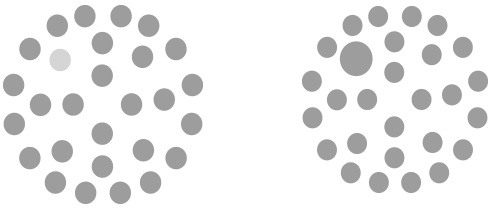

Fig. 1 The display of the visual search with luminance target (left) and size traget (right).

2.3.2 Movement task apparatus: The movement cues (Fig. 2) were presented in the middle of the computer screen. The respective movement was to be executed on one of the three different paper cups, placed 25 cm away and below the computer screen (see Fig. 3). The cups differed in size and luminance: a small white, 5 cm in diameter in the middle point; a middle grey, 6.5 cm in diameter in the middle point; and a large dark grey cup, 8 cm in diameter in the middle point. They were all equal in height and weight. Participants began the movement subsequent to the visual search task (see the Procedure section), upon presentation of a go-signal. The go-signal was a yellow asterisk of 0.6˚in diameter. It was presented 4.5˚, 11.3˚, or 17.7˚from the left border of the screen and it indicated the cup, which was positioned directly below the location of the asterisk. To circumvent learning of location for each of the cup, cup positions were changed after each block: if the small cup was placed on, for example, the leftmost position in Block 1, it was placed on the, for example, rightmost location for Block 2. These locations were randomly selected across blocks.

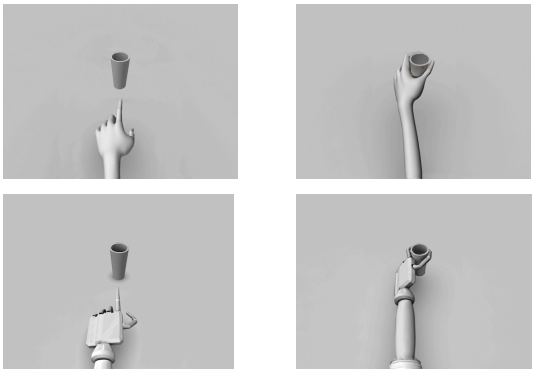

Fig. 2 Top: Human cartoon hand signaling pointing movement (left) and grasping movement (right). Bottom: Robot cartoon hand signaling pointing movement (left) and grasping movement (right).

2.4 Procedure: All participants were seated in a quiet and dimly lit room with response mouse positioned under their dominant hand and placed on the lap (see Fig. 3).

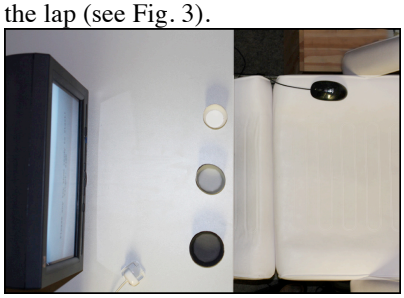

Fig. 3 Experimental setup inside the chamber, view from the top. The three cups are visible as circular objects at the edge of the table.

2.4.1 Trial sequence: At the beginning of each trial a 300 ms fixation cross (x) was displayed in the center of the screen. Subsequently, the movement cue was presented for 800 ms. Subsequent to another fixation cross (200 ms) a visual search display was presented for 100 ms (see Fig. 4). The visual search display was followed by a blank screen, during which participants responded to the visual search display (target present vs. target absent). Reaction time was measured from the offset of the visual response to the moment of the key press. Upon the visual search response, and another fixation cross (400 ms), the go-signal for the movement execution was presented for 300 ms. Participants then executed the prepared movement. Movements were registered by the experimenter with the use of a web camera (Microsoft LifeCam VX 800) and a computer mouse. The next trial began subsequent to the experimenter's registration of the performed movement.

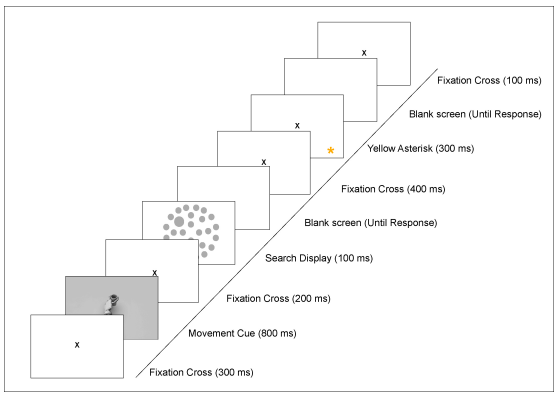

Fig. 4 A trial sequence of the present experiment.

2.4.2 Experimental protocol: Participants were instructed to respond as fast and accurate as possible in the search task. In the movement task only accuracy was stressed. Participants were provided with feedback concerning their performance after each block. Visual search target type (luminance vs. size) was blocked and the order of blocks was counterbalanced across participants. That is, half of participants performed a size

detection task for 240 trials, and then the luminance detection task for another 240 trials; while the other half had the reverse order of visual search target type.

The two movements (grasping or pointing) were randomized within blocks, across individual trials. The cue types (robot vs. cartoon) were presented in two separate experimental sessions on two separate days. Altogether, each participant took part in three sessions on three separate days. The first session consisted in practicing only the movement task (15-30 min), so that the subsequent experimental sessions involving two tasks would be easier to perform. During the first (practice) session participants practiced the grasping and pointing movements in 5 blocks. In four blocks, 24 trials each, only pointing or only grasping was performed. In one block, both movements (60 trials) were performed in a randomized order. The two experimental sessions consisted in two tasks: the visual search and the movement task, as described above and depicted in Fig. 4. In one of the sessions participants were presented only with robot movement cues and in the other session only with human cartoon movement cues. Each of the experimental sessions consisted in 2 practice blocks (one with movement only and one with both tasks) and 8 experiment blocks, 60 trials each. The visual search target type (size vs. luminance) changed after 4 blocks (240 trials).

2.5 Data Analysis

Error rates were computed for each participant in both the search task and the movement task. Two types of responses were treated as errors in the search task: (i) a positive response (response "target present") in target absent trials – the so-called "false alarm"; and (ii) a negative response (response "target absent") in target present trials – the so-called "miss". In the movement task, an error was committed when a participant grasped an object instead of pointing to it; or vice versa. Data of three participants whose error rates in any of the search tasks were above 15% were excluded from the analyses (mean error rates for the other participants: 7%). We considered participants whose error rates were higher than twice the mean for other participants as outliers; and decided to exclude their data from analysis due to that such high error rates might indicate not sufficient focus on the task, or not sufficient comprehension of the required task (typically, the error rates in experiments with this type of paradigm are not higher than 10% on average and are lower than 15% for individual participants, see [3, 12-14]).

Prior to RT analysis in the search task, errors in any of the two tasks were excluded. RTs longer than 1100 ms and shorter than 50 ms were treated as errors and excluded as well. Data of one participant were excluded due to abnormally long RTs in the luminance detection task (M = 791 ms) while other participants' mean RT in this condition was 465 ms, with the range between 357 ms to 635 ms (individual mean RTs). Similarly to the exclusion criteria based on error rates, we considered the data set of this participant as an outlier – presumably being due to the participant not being sufficiently focused on the task. From the remaining data set (sample: n = 16), median RTs and mean error rates were calculated and subject to an analysis of variance (ANOVA) with the within-subject factors of cue type (robot vs. cartoon), task type (size vs. luminance), movement type (grasping vs. pointing), display type (target vs. blank). Error rate analyses in the search task were conducted on correct movement trials.

3 Results

3.1 Reaction times

The analysis performed on the median RT data revealed a main effect of display type, $F(1,15) = 9.78$, $p = .007$, $\eta_p^2 = .395$ showing faster RTs to target trials (M = 358 ms) as compared to blank trials (M = 397 ms). Most importantly, the interaction of movement type and task type was significant, $F(1,15) = 9.24$, $p = .008$, $\eta_p^2 = .381$ (see Fig. 5), indicating the expected congruency effects. This interaction was further tested with planned comparisons, which revealed congruency effects for both target types: RTs in the size detection task, were faster for the grasping (congruent) condition (M = 370 ms) relative to pointing (M = 376 ms), $t(15) = 2.37$, $p = .016$, one-tailed; while RTs in in luminance detection were faster for the pointing condition (M = 379 ms) relative to grasping (M = 384 ms), $t(15) = 1.88$, $p = .039$, one-tailed[1].

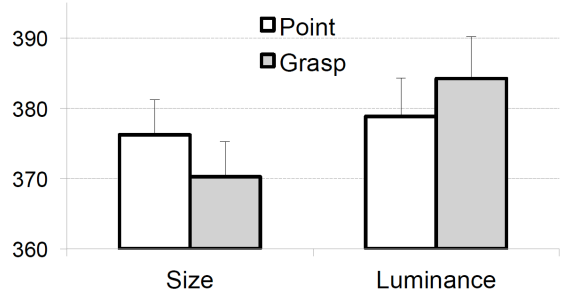

Fig. 5 Average median reaction times (RTs) as a function of task type (luminance vs. size) and movement type (pointing vs. grasping). The differences between the movement types for each

[1]Note that when the two left-handed participants were excluded from analysis, the pattern of results remained the same, with significant interaction of movement type and task type, $F(1,13) = 7.34$, $p = .018$, $\eta_p^2 = .361$; with faster RTs in the size taks for grasping (M = 376 ms), relative to pointing (M = 381 ms), $t(13) = 1.91$, $p = .039$, one-tailed; and faster RTs in the luminance task for pointing (M = 380 ms) relative to grasping (M = 386 ms), $t(13) = 1.92$, $p = .035$, one-tailed. The interaction with type of cue (human vs. robot) was not significant, $F(1,13) = .17$, $p = .68$, indicating that the congruency effects were similar for both human and robot cartoon hands. Thus, even though two participants were naturally left-handed; their data did not affect the pattern of results.

of the visual search tasks are the congruency effects. Error bars represent standard errors of the mean adjusted for within-participants designs, calculated according to the procedure described in [22].

Importantly for the purposes of this experiment, the interaction between movement type and task type (congruency effects) did not depend on cue type (robot vs. human), $p > .98$ (see Fig. 6), suggesting that congruency effects were elicited both by human and robot hands.

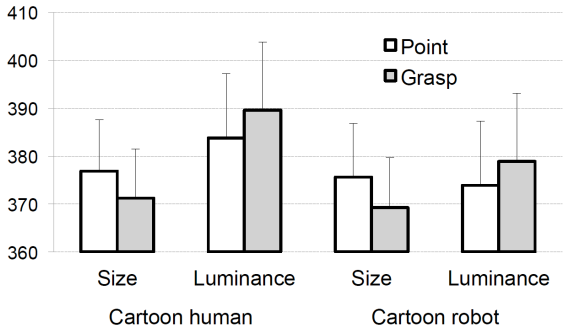

Fig. 6 Average median reaction times (RTs) as a function of movement type (point vs, grasp), task type (size vs. luminance) and two types of cues (cartoon or robot). Error bars represent standard errors of the mean adjusted for within-participants designs, calculated according to the procedure described in [22].

3.2 Error Rates

Analogous analysis on mean error rates showed only the main effect of display type, $F(1,14) = 5.87$, $p = .029$, $\eta_p^2 = .296$ with less errors for blank trials (M = 4.3%) than for target trials (M = 7.9%), revealing a reverse pattern than in the RT data. The analysis yielded also a main effect of task type $F(1,14) = 5.16$, $p = .039$, $\eta_p^2 = .269$ with less errors in the size task (M = 4.8%) as compared to the luminance task (7.4%), suggesting that the size task was slightly easier than the luminance task. Moreover, the main effect of movement type reached the level of significance, $F(1,14) = 14.46$, $p = .002$, $\eta_p^2 = .508$, showing better performance in the pointing condition (M = 5.2%) as compared to grasping (M = 7%) suggesting that pointing might have been cognitively less demanding than grasping. These effects are further discussed in the Discussion section. The interaction between movement type and task type (congruency effect) was not significant, $p > .14$.

4 Discussion

The aim of this experiment was to test whether observing robot hands or cartoon human hands performing two types of movement (grasping or pointing) would elicit similar action representations in the human brain as when viewing human hands in action. This was tested indirectly through examining the impact that action representations have on perceptual processes. We used a modified paradigm of Wykowska and colleagues [3], in which participants are typically required to perform a perceptual task (visual search for a target defined by either size or luminance dimension) and a movement task (grasping or pointing). The tasks are unrelated both motorically and perceptually, as the visual search objects are presented on the computer screen (and participants respond with a key press on a computer mouse with their dominant hand) while the movement task objects are placed below the computer screen (and participants respond by either pointing or grasping one of the objects with the other hand). Importantly, in the present study the to-be-performed grasping or pointing movement was signaled by a picture cue representing either a human-like cartoon hand or a robot hand, in contrast to previous studies in which the cue depicted human hands. We reasoned that if observing a robot hand elicits a representation of action similar to when another human is observed, the action representation should have a similar impact on perception as in case of representation of human actions, and should be manifested by the so-called action-perception congruency effects observed in previous research for human actions [3, 12-14]. Action-perception congruency effects should be observed in the form of better performance (faster RTs and/or lower error rates) for visual search targets in the action-congruent conditions, relative to action-incongruent conditions. In the case of the present paradigm, grasping was the congruent action with the size targets while pointing was congruent with luminance targets.

Our present results indeed showed better performance for target detection in the action congruent conditions. That is, size targets were detected faster when participants were planning to grasp, relative to point; while luminance targets were detected faster when participants were planning to point, relative to grasp. Most importantly, these effects did not depend on what type of cue signaled the movement, and were observed for both the human cartoon hands as well as the robot hands. This suggests that when humans observe humanoid robot hands in action (robot hands that are very similar to human hands both in morphology as well as implied functionality), similar representation of an action is elicited, as when another human's action is viewed.

This is particularly striking, as not every type of action cue elicits an action representation that is effective in influencing perceptual processing. For example, Wykowska et al. [13] showed that word cues do not produce congruency effects, in contrast to picture cues. The authors argued that word cues are not as effective as picture cues in triggering the intentional weighting mechanism. This might be due to that word cues are less likely to evoke a sort of an action template for the required action. Interestingly, this does not result in poorer action performance itself (based presumably on offline

representations and invariant characteristics of an action), but rather influences perceptual selection processes to a lesser degree than in case of pictorial cues (an influence related to the intentional weighting mechanism for online action control).

Interestingly, observing robot hands in action is equally effective in evoking an action representation that has an impact on perceptual processes as observing human hands. Therefore, it seems that observing humanoid robot actions might result in efficient simulation of the observed actions, thereby making the perceptual system prepared for fast and efficient delivery of relevant perceptual characteristics for online action control. The present findings are in line with the media equation theory [19] in that they suggest an automatic and natural response to an artificial agent (perhaps even implying personification). Our results are also in accordance with a study by Oberman et al. [23] who found that observing both human and robot actions activated the mirror neurons in the human brain. The mirror neuron activation facilitated reproduction of the same observed action. Oztop et al. [24] also found that both humans and humanoid robots elicited interference effects when being observed simultaneously with execution of movements that could be either congruent or incongruent with the observed ones. Similarly, Press et al. [20] observed that humans simulate not only human but also robot actions (although to a somewhat lesser degree), as indicated by stimulus-response compatibility effects (or visuomotor priming) for both human and robot stimuli when robot hands were similar to human hands. Interestingly when the robot stimuli were made perceptually different from human hands [25], the simulation mechanism was activated to a lesser extent than in the case of human hands.

This suggests that it is important to note that the stimuli used in the present design were static pictures of robot hands with very human-like morphology. It remains to be answered whether a similar pattern of results would be observed if a video of robot movement was presented to participants; and if the shape of the hand was less similar to human hands. These questions will be addressed in future research (see below).

Crucially, the present study extends previous findings by that we examined the perceptual consequences of simulation and not the visuomotor priming effects per se. That is, paradigms used in [20, 24-25] targeted at the processes related to movement production itself. Our study, in contrast, was designed to pinpoint a more general mechanism of *perceptual* selection: selection with respect to action planning. We aimed at measuring if processing perceptual dimensions is tuned to action planning when a robotic hand signals an action. Importantly for our design, the perceptual task and the motor task were completely unrelated (in contrast to the previous studies [20, 24-25]) and therefore, the effects are presumably due to overlap between the action and perception domains at the representational level in the brain.

In addition to the effects of major theoretical interest, the present analyses revealed also that participants were faster in responding to target present trials as compared to target absent trials, which is a common finding in the visual search literature [26] indicating different processing modes for situations when a signal is present in the visual field as compared to being absent [27, 28]. Error rates however, were larger for target absent trials, as compared to target present trials, suggesting some degree of speed-accuracy tradeoff related to target detection. Importantly, however, no speed-accuracy tradeoff was observed for the effects of interest (the congruency effects). Error rates depended also on the type of movement revealing better performance in the pointing condition as compared to grasping; and on the type of task: size targets were detected with lower error rates than luminance targets. This suggests that performance in the visual task depended on the cognitive load with simpler conditions (pointing movement) yielding better performance than more demanding conditions (grasping). Lower error rates for size targets compared to luminance targets suggest that size might have been a slightly more salient dimension than luminance. Neither of these effects had an impact on the effects of interest, however, as the interaction of movement type and task type did not reach the level of significance.

5 Implications for social robotics

The experimental design presented in this paper can serve as means for implicit measures of understanding robot actions (see [29] for an overview on measurements in social robotics contexts). If we assume that simulating observed actions recruits similar neuronal mechanisms as those involved in actual action execution; and that action simulation is one of the fundamental mechanisms of action understanding, then the congruency effects that can be measured with a paradigm presented in this paper should be good (and implicit) indicators of whether the observed action is properly mapped (and thereby understood) to the observers' representational system. This, in turn, can be an indirect measure of acceptance of robots as actual social partners, and not only simple automata. To give an example, it seems unlikely that humans simulate the workings of a machine such as, for example, a simple printer, which draws a sheet of paper, plots the required text/images and produces an output in form of a printed document. It seems unlikely that the human cognitive system would map such actions to their own action repertoire thereby generating action representations that would be based on the same neural architecture

as action planning itself, which would then affect perceptual processing (congruency effects). In contrast, when observing an entity that is more human-like and behaves in a human-like manner, humans are definitely more likely to map the observed actions to representations of their own actions that are generated in their cognitive system.

This yields an important and interesting question: at which point actions of a machine become possible to be mapped to one's own action representations; and what are the parameters that are crucial for eliciting simulation-based representations of the observed actions. Previous research [30, 31] has shown that our own motor repertoire limits the extent to which we simulate actions of others. For example, when expert ballet dancers observed other ballet experts dancing, activity in the motor-related brain regions depended on whether the observed dancers were of the same gender or not, independent of familiarity of the particular dance and movement patterns (the observers and dancers practiced together). Therefore, one of the important issues that social robotics needs to address is how to design robots, which actions elicit in human brains appropriate action representations and fundamental mechanisms underlying proper action understanding. In this paper, we propose that this is one of the crucial aspects of treating the robots we interact with as social companions and not only simple automata.

6 Future directions

In the present study, we examined whether pictures of a humanoid robot hand in action would elicit an action representation in the human brain that can be mapped to observer's own action repertoire. This is the first step to systematically examine particular parameters of the robot morphology and/or movement dynamics that can potentially have an impact on efficiency with which robots in action evoke appropriate action representations, allowing action understanding. In the present study, we used static pictures of a robot hand resembling to a large extent a human hand - both in morphology and in implied functionality. Therefore, the present findings should be considered only in the context of the stimuli used. Future research will test (i) various degrees of similarity of robot morphology to human morphology; and (ii) various degrees of similarity in movement dynamics to human movement dynamics - in order to ultimately answer the question of how those two dimensions impact the way humans represent robot actions; and in order to be able to examine if the findings are generalizable to various robot shapes and/or movement dynamics.

7 Conclusions

Results of the present work suggest that static pictures of humanoid robot hands, which - to a large extent - resemble human hands are capable of evoking appropriate action representations in the human brain that allow for mapping the observed actions onto observer's own action repertoire. If this is the case for dynamic robot actions with different morphology remains to be answered. Importantly, however, this approach offers a unique method of systematic examination in an implicit manner human's ability to understand robot actions; and proposes to apply the present paradigm to systematic testing of the impact that various parameters of robot's morphology and/or movement dynamics have on the way humans represent robot actions. This in turn, should help in designing robots that are to act and interact with humans in a social and intuitive for humans manner. Ultimately, this approach might provide an answer to one of the crucial questions in social robotics: how to construct robots, which would be treated as social companions and not only as simple automata.

Acknowledgements This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) - grant awarded to AW (WY-122/1-1).

8 References

[1] Greenwald A, Sensory feedback mechanisms in performance control: With special reference to the ideomotor mechanism. Psych Rev, 77: 73-99 (1970).

[2] Hommel B, Müsseler J, Aschersleben G, Prinz W The theory of event coding (TEC): A framework for perception and action planning. Behav Brain Sci 24:849-878 (2001).

[3] Wykowska A, Schubö A, Hommel. B, How you move is what you see: Action planning biases selection in visual search. J Exp Psychol: Human 35:1755-1769 (2009).

[4] Baron-Cohen S, Mindblindness: an essay on autism and theory of mind. Boston: MIT Press/ Bradford Books (2006).

[5] Frith CD, Frith U, How we predict what other people are going to do. Brain Res 1079: 36-46 (2006).

[6] Decety D, Grèzes J, Neural mechanisms subserving the perception of human actions. Trends Cogn Sci 3: 172-178 (1999).

[7] Rizzolatti G, Fogassi L, Gallese V, Neurophysiological mechanisms underlyingthe understanding and imitation of action. Nat Rev Neurosci 2: 661-670 (2001).

[8] Schilbach L, Timmermans B et al., Toward a second-person neuroscience. Behav Brain Sci 36:393-462 (2013).

[9] Kilner JM, Friston KJ, Frith CD, Predicitve coding: an account of the mirror neurons system. Cogn Process 8:159-166 (2007).

[10] Wolpert DM, Ghahramani Z, Computational principles of movement neuroscience. Nat Neurosci 3: 1212-1217 (2000).

[11] Hommel B, Grounding attention in action control: The intentional control of selection, in E□ortless attention: A new perspective in the cognitive science of attention and action, ed. BJ Bruya (Cambridge, MA: MIT Press), 121-140 (2010).

[12] Wykowska A, Hommel B, Schubö A, Action-induced effects on perception depend neither on element-level nor on set-level similarity between stimulus and response sets. Atten Percept Psycho,73:1034-1041 (2011).

[13] Wykowska A, Hommel B, Schubö A, Imaging when acting: picture but not word cues induce action-related biases of visual attention. Front Psychol 3:388 (2012).

[14] Wykowska A, Schubö A, Action intentions modulate allocation of visual attention: electrophysiological evidence. Frontiers in Psychology, 3:379 (2012).

[15] Mori M, Bukimi no tani The uncanny valley (KF MacDorman, T Minato, Trans.). Energy 7:3335. (Originally in Japanese) (1970).

[16] Saygin AP, Chaminade T, Ishiguro H, Driver J, Frith C, The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive Affective Neuroscience*, 7:413-22 (2012).
[17] Moore RK, A Bayesian Explanation of the 'Uncanny Valley' Effect and Related Psychological Phenomena. *Sci Rep* 2 (2012). doi:10.1038/srep00864.
[18] Freier, NG, Kahn Jr, PH, The Fast-Paced Change of Children's Technological Environments. *Children Youth and Environments*, *19*, 1-11 (2009).
[19] Reeves B, Nass C, The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge University Press (1996).
[20] Press C, Bird G, Flach R, Heyes C, Robotic movement elicits automatic imitation, Cognitive Brain Research, 25, 632-640 (2005).
[21] Anderson SJ, Yamagishi N, Spatial localization of colour and luminance stimuli in human peripheral vision. Vis Res 40: 759771 (2000).
[22] Cousineau D, Confidence intervals in within-subject designs: A simpler solution to Loftus & Masson's method. Tutorials in Quantitative Methods for Psychology, 1, 42–45 (2005).
[23] Oberman LM, McCmeery JP, Ramachandran VS, Pineda JA, 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 (2007).
[24] Oztop E, Franklin DW, Chaminade T, Cheng G, Humanhumanoid interaction: Is a humanoid robot perceived as a human? Int J Hum Robot 2:537559 (2005).
[25] Press C, Gillmeister H, Heyes C, Bottom-up, not top-down modualtion of imitation by human and robotic models, European J of Neurosci, 1-5 (2006).
[26] Chun MM, Wolfe JM, Just Say No: How Are Visual Searches Terminated When There Is No Target Present? Cognitive Psychol 30:39-78 (1996).
[27] Schubö A, Schröger E, Meinecke C Texture segmentation and visual search for pop-out targets. An ERP study. Cognit Brain Res 21:317-334 (2004).
[28] Schubö A, Wykowska A, Müller HJ Detecting pop-out targets in contexts of varying homogeneity: Investigating homogeneity coding with event-related brain potentials (ERPs). Brain Res 1138:136-147 (2007).
[29] Brayda L, Chellali R, Measuring Human-Robot Interactions, Editorial of Special Issue, International Journal of Social Robotics, Springer. DOI: 10.1007/s12369-012-0150-(2012)
[30] Schütz-Bosbach S, Prinz W, Perceptual resonance: action-induced modulation of perception. Trends Cognit Sci, 11: 349-355 (2007).
[31] Bosbach S, Cole J, Prinz W, Knoblich G, Inferring anothers expectation from action: the role of peripheral sensation. Nat Neurosci 8:12951297 (2005).

Biographical notes

Agnieszka Wykowska is a senior research associate and a lecturer (Privatdozent) at the Department of Psychology, Ludwig-Maximilians-Universität (LMU) München, Germany and at the Institute for Cognitive Systems, Technische Universität München (TUM), Germany. She obtained PhD in Psychology from the LMU in 2008. Agnieszka Wykowska's background is Cognitive Neuroscience (M.Sc. in Neuro-cognitive Psychology from the LMU, Munich) and Philosophy (M.A. in Philosophy, from the Jagiellonian University Krakow, Poland). Since 2006, she has been employed in the Department of Psychology, LMU, being involved in a Cluster of Excellence CoTeSys (Cognition Technical Systems) - an interdisciplinary project aiming at developing robots with cognitive capabilities. Since September 2013, she has also been working at the Institute for Cognitive Systems, Technische Universität München. Her research interests include visual attention and perception, action-perception links, action planning, social attention and joint action; as well as human-human and human-robot interaction in the context of social robotics. In her research, she uses psychophysics and EEG/ERP methodology.

Ryad Chellali is a senior scientist at the Department Pattern Analysis and Computer Vision (PAVIS), Istituto Italiano di Tecnologia. He obtained his PhD in Robotics from University of Paris in 1993 and his Dr. Sc from Univesrity of Nantes (France) in 2005. His main research interests include robotics, human-robots interactions, human behavior analysis (social signal processing and a□ective computing). Telepresence virtual and augmented realities, are also keywords of his activity. He served as Junior Researcher in 1992 at the French Institute of Transports (INRETS). From 1993 to 1995 he was assistant professor at University of Paris. From 1995 to 2006, he joined Ecole des Mines de Nantes/CNRS (France), heading the automatic control chair. He joined IIT in 2006 as a senior scientist, where he created the Human-Robots Mediated Interactions Lab. Ryad Chellali co-authored more than 100 papers. In 2000 and 2005 the French Government awarded him for the creation of innovative technologies comapnies.

Md. Mamun Al-Amin graduated with Master's Degree in Neuro-cognitive Psychology from the Ludwig-Maximilians-Universität, Munich, Germany. Besides his Master study, he worked at the psychiatric hospital of the Ludwig-Maximilians-Universität between 2011 and 2012. Before coming to Munich, Md. Mamun Al-Amin was an employee of the General Pharmaceuticals in Bangladesh. He joined the Department of Pharmaceutical Sciences, North South University (Dhaka, Bangladesh) in January 2013. He is currently employed as a Lecturer and practices neuro-cognitive research besides lecturing. He is interested in Neuro-cognitive science. His research combines neuroscience and pharmacology (drug action) to explore the effect of new and old chemical entities on cognitive function.

Hermann Müller studied Psychology at the University of Würzburg, Germany. He then went to the University of Durham, UK, to carry out Ph.D. research on visuo-spatial orienting of attention; Müller & Rabbitt (JEP:HPP, 1989; approx. 1.000 citations) is one of the several high-impact papers coming out of this work. There followed a period of Post-doc work at the Universities of Manchester and London (Birkbeck College), UK, collaborating with Patrick M.A. Rabbitt and Glyn W. Humphreys, respectively, with visual search becoming a more prominent research issue. In 1992, Hermann Müller was appointed Lecturer of Psychology at Birkbeck College (University of London), quickly rising to Senior Lecturer (1995) and then Reader of Psychology at Birkbeck (1997). In 1997, he was appointed Professor of Experimental Psychology at Leipzig University, Germany. In 2000, he moved to the Ludwig-Maximilians University of Munich to take up the Chair of General and Experimental Psychology. In Munich, he built a highly active group devoted to the study of selective attention and action control, pursuing a neuro-cognitive research approach. During his career, he has published over 200 original papers in major journals of Experimental Psychology and the Cognitive Neurosciences, and he has initiated and organized a prestigious series of international symposia devoted to the study of Visual Search and Selective Attention (VSSA). In 2010, he was awarded the Wilhem Wundt price of the Wilhelm Wundt Society for "excellent achievements in fundamental psychological research".