CrossMark

Humans are Well Tuned to Detecting Agents Among Non-agents: Examining the Sensitivity of Human Perception to Behavioral Characteristics of Intentional Systems

Agnieszka Wykowska[1,2] · **Jasmin Kajopoulos**[1,3] · **Miguel Obando-Leitón**[4] · **Sushil Singh Chauhan**[5] · **John-John Cabibihan**[6] · **Gordon Cheng**[2]

Abstract For efficient social interactions, humans have developed means to predict and understand others' behavior often with reference to intentions and desires. To infer others' intentions, however, one must assume that the other is an agent with a mind and mental states. With two experiments, this study examined if the human perceptual system is sensitive to detecting human agents, based on only subtle behavioral cues. Participants observed robots, which performed pointing gestures interchangeably to the left or right with one of their two arms. Onset times of the pointing movements could have been pre-programmed, human-controlled (Experiment 1), or modeled after a human behavior (Experiment 2). The task was to determine if the observed behavior was controlled by a human or by a computer program, without any information about what parameters of behavior this judgment should be based on. Results showed that participants were able to detect human behavior above chance in both experiments. Moreover, participants were asked to discriminate a letter (F/T) presented on the left or the right side of a screen. The letter could have been either validly cued by the robot (location that the robot pointed to coincided with the location of the letter) or invalidly cued (the robot pointed to the opposite location than the letter was presented). In this cueing task, target discrimination was better for the valid versus invalid conditions in Experiment 1 where a human face was presented centrally on a screen throughout the experiment. This effect was not significant in Experiment 2 where participants were exposed only to a robotic face. In sum, present results show that the human perceptual system is sensitive to subtleties of human behavior. Attending to where others attend, however, is modulated not only by adopting the Intentional Stance but also by the way participants interpret the observed stimuli.

✉ Agnieszka Wykowska
  agnieszka.wykowska@psy.lmu.de

  Jasmin Kajopoulos
  jasmin.kajopoulos@campus.lmu.de

  Miguel Obando-Leitón
  miguel.obando@tum.de

  Sushil Singh Chauhan
  sushil.chauhan@nus.edu.sg

  John-John Cabibihan
  john.cabibihan@qu.edu.qa

  Gordon Cheng
  gordon@tum.de

1 General and Experimental Psychology Unit, Department 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 Neuro-cognitive Psychology Master Program, Department of Psychology, Ludwig-Maximilians-Universität, Leopoldstr. 13, 80802 Munich, Germany

4 Graduate School of Systemic Neurosciences, Ludwig-Maximilians-Universität, Großhaderner Str. 2, 82152 Planegg-Martinsried, Germany

5 Singapore Institute for Neurotechnology (SINAPSE), National University of Singapore, 28 Medical Dr. #05-COR, Singapore 117456, Singapore

6 Department of Mechanical and Industrial Engineering, Qatar University, Doha 2713, Qatar

Keywords Turing test · Agency · Human–robot interaction · Pointing · Spatial cueing

1 Introduction

In everyday lives, humans are almost always embedded in a social context, which involves various types of interac-

🖉 Springer

tions. This in turn necessitates understanding and predicting others' behavior. The human brain has developed mechanisms, which use various hints that others provide as cues for explaining their behavior: facial expressions, gaze behavior or gestures. Several authors, e.g., [1,2] postulate that based on such hints, humans infer mental states underlying the observed behavior, which allows for explaining the behavior. For example, if I see someone seated at a table and gazing in the direction of a glass with orange juice, I might infer—thanks to gaze being a social cue—that the observed person is intending to grasp the glass, presumably with the desire of drinking the orange juice. I, therefore, predict that the person will next grasp the cup, lift it and bring it to his/her lips. This sort of reasoning—the so-called *mentalizing* [2] or *Theory of Mind* [1]—allows for explaining behavior and predicting subsequent action steps.

However, it is important to note that before one can activate the process of mentalizing, i.e., attributing particular mental states to the observed behavior, one needs to assume that the observed entity is capable of having mental states. In other words, one needs to treat the observed entity as a rational agent with a mind. Dennett [3] conceptualizes this as adopting the *Intentional Stance*. According to Dennett, Intentional Stance is the best strategy for explaining behavior of intentional systems. That is, when I want to explain and/or predict the behavior of another human, I will be much more successful when I refer to his/her mental states which might underlie the observed behavior, rather than, e.g., low-level physical states. For example, if I want to predict what a soccer player will do in the next move, I will be more successful if I assume the soccer player is an intentional agent and refer to his mental states such as "desire to win the game", "intention to score a goal", etc. rather than when I refer to the current state of particle dynamics in his body and environment. In contrast, when I want to explain a certain chemical reaction in a laboratory, I better refer to the physical properties of particles, rather than assume their intentionality and refer to their mental states. Dennett distinguishes three stances: the physical, the design and the Intentional Stance; each being best suited to the level of description at which one operates in a given context.

Adopting the Intentional Stance has been shown to activate specific brain regions [4,5] and, importantly, modulate mechanisms of social cognition, such as joint attention (i.e., attending to where others attend) [6,7]. That is, participants were more ready to engage in joint attention (i.e., attended where the observed entity gazed) when they believed that the observed entity's behavior was controlled by a human (Intentional Stance adopted), rather than by a computer program. This was also reflected in a modulation of an early component of the EEG signal [7], namely the P1 occurring around 100 ms post stimulus onset, indicating that the way a stimulus is processed at the sensory level is affected by atten-

tional processes, which in turn are influenced by higher-order cognition.

Adopting the Intentional Stance seems to be a plausible adaptive mechanism, as in the course of interacting with others, one needs to know and understand if the observed behavior results from operations of the mind (and therefore can potentially carry intentional and socially informative content) or is just a consequence of non-intentional behavior. Tomasello [8] accounts for this with his distinction about two types of intentions being communicated by social gestures: a referential intention (where attention is oriented) and a social intention (what is the reason to direct an interaction partner's attention to that location). Obviously, social intention is missing in non-intentional systems, as an entity without a mind cannot have intentional reasons to direct others' attention to a location of interest. Therefore, it seems indeed very important to know whether the observed entity is an agent with a mind, and thus, whether the entity's behavior provides some social meaningful content. To give an example, imagine that you are driving a car and you observe that a car behind you started blinking with its right blinker. You know that the car behind you is being operated by your friend with whom you are going on vacation, and that this way your friend signals to you that she would want to turn right in the next crossing. Your attention is therefore oriented towards the right—and towards the location in which the next exit would appear. In a different scenario, you might imagine that the car behind you is actually pulled by your car on a rope, because it broke down, and the electrical circuit controlling the blinkers is broken. The blinkers randomly turn on or off. In this situation, you will probably not orient your attention towards the direction indicated by the blinker. In fact, you might learn to ignore the blinking completely. This example shows that assuming intentionality is extremely important for the way humans interpret behavior and for the way they react upon the observed behavior.

1.1 Aim of the Present Study

In the previous studies [4–7] adopting the Intentional Stance was the result of explicit instructions that participants obtained from the experimenter. One of the key questions that remains to be answered is: under what conditions humans spontaneously adopt the Intentional Stance (attribute mind) towards other agents. In other words, what are the specific characteristics of behavior of an agent with a mind, and how much is the human's perceptual system sensitive to them.

We aimed at addressing this question with the use of a type of "Turing test". The concept of the Turing test was first proposed [9] as a criterion for intelligence. According to Turing [9], attribution of intelligence to any entity arises from observed behavioral cues; and this should constitute a sufficient criterion for an intelligent mind. More

recently, Pfeiffer et al. [10] developed an actual experiment in which the logic of the Turing test was used. In their study, participants observed an avatar and were told that their interaction partner can be a computer or a human. Participants' task was to discriminate human behavior from a computer program, although in reality, the avatar's behavior was always controlled by a computer program. The results showed that humans attribute humanness to the avatar based on assumptions they have concerning other humans' behavioral patterns. The study of Pfeiffer et al. [10], however, did not test whether humans are sensitive to other humans' behavior, as the authors always used a computer program that controlled the behavior of the avatar.

Our study, in contrast, aimed at answering the question of whether humans are sensitive to other human's behavior and if so, what are the parameters (even if very subtle) that provide hints for detecting other humans through only observation of behavior.

1.2 Design

To examine human ability to discriminate between behavior of an agent from a mechanistic behavior, we introduced a paradigm in two experiments in which participants interacted with a robotic platform that had arms pointing in various directions. In the two experiments, we used two different robots to test the ability of participants to discriminate human-like behavior. In the first experiment, two robotic arms were placed in front of a static picture of a human face, while in the second experiment we used a NAO robot (Aldebaran Robotics) with a fully robotic yet humanoid appearance. A previous paper [11] investigated the effect of the agent's nature, and used images of a human face as well as of a NAO robot for a gaze cueing task. They reproduced an effect of validity on reaction times (RTs) for both agents, as well as increased RTs when using the robotic face, which they interpreted as a difficulty to disengage attention from a novel stimulus. In our experiment, we investigated the effect of the nature of agency (human vs. mechanistic) within the same type of appearance (Human in Experiment 1; Robot in Experiment 2).

In the present design, in some blocks, the onset time of the arm movement was controlled by a computer program, and in some other blocks it was controlled by an experimenter seated in a separate room (Experiment 1) or modeled after human behavior (Experiment 2). Participants performed a "Turing test" by determining (at the end of the block) whether they had interacted with a human or with a computer-controlled interface. The only hint participants had concerning whether the arms were controlled by a human or by a program was variability in movement onset times in the human condition. Importantly, the arm movements themselves and all other factors were exactly identical across conditions. Participants

were not explicitly instructed regarding the hint they could use to perform this judgment, which allowed testing how sensitive the human perceptual system is to subtle human-like behavioral characteristics.

Apart from the "Turing test", participants performed a task in which they discriminated a letter "T" or "F". Before the letter appeared, the robot arm would point either in the direction of where the letter would then be presented (valid trials) or in the opposite direction (invalid trials). The procedure, therefore, followed the logic of a gaze-cueing paradigm ([6,7,12,13], see also [14] for a review). In gaze cueing paradigms, a face stimulus is typically presented in the middle of a computer screen. In the course of a trial sequence, the face shifts gaze direction towards one of the sides of the visual field. Subsequently, a target is presented (e.g., a letter), and participants' task is to detect, localize or discriminate the letter. It is assumed that if attentional focus follows the direction of the gazer's eyes, target-related performance should be better for the conditions in which the gazer looked towards the side at which a target subsequently appeared (validly cued condition), as compared to when the gazer looked at the opposite side (invalidly cued condition). Results typically show this pattern [6,7,12,13].

The present paradigm followed a similar logic with the only difference that instead of gaze direction, we used pointing movements. Pointing gestures are strong behavioral cues in human interactions, showing readiness to interact with another person [15], emerge in early developmental stages as a declarative gesture in order to share attention [16,17] and are a stronger cue than lexical information in early childhood [18]. Even when pointing gestures are made by a telepresence robot, human observers can better comprehend spatial information provided by pointing gestures and verbal instructions, as compared to when instructions are only provided verbally [19,20]. Taken together, pointing gestures are socially as important as gaze/head direction [21]; with the crucial difference being that pointing inherently presupposes an intention to direct others' attention to some location, while gaze can reflect a reflexive process of attention being oriented to a location upon a salient environmental signal. Therefore, we hypothesized that pointing gestures should be socially more involving than gaze direction, and hence were better suited to examining mechanisms of social cognition in a more naturalistic social interaction scenario. We expected participants to covertly attend the location where the robot pointed and therefore discriminate target letter better (faster or with higher accuracy) in the valid, relative to the invalid condition. Moreover, we reasoned that participants might be more likely to covertly attend where the robot points (show larger validity effects) when the robot behavior would be perceived as human-controlled, relative to when it would be perceived as pre-programmed, because in the pre-programmed condition participants would not adopt the Intentional Stance towards

the robot behavior, and thus they would not assume any intentions and social content involved in the pointing gestures.

2 Experiment 1

2.1 Methods

2.1.1 Participants

Twenty-four participants (Mean age 22.21, SD 4.94; four men) took part in this experiment for an honorarium. All participants were healthy volunteers and had normal or corrected-to-normal vision. Three participants were left-handed. The experiment was conducted with the full understanding and written consent of each participant. Data of two participants had to be discarded due to technical problems during data acquisition, which resulted in insufficient amount of recorded data points.

2.1.2 Ethics Statement

The experiments were conducted at the Social Robotics Laboratory, National University of Singapore. All participants were healthy and adult. The experimental procedures consisted of purely behavioral data collection (RTs and error rates), and filling out a questionnaire. The procedures did not include invasive or potentially dangerous methods and were in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Data were stored and analyzed anonymously. Participants gave written consent and received monetary compensation for participating. These standard procedures are approved by the ethics committee of the Department of Psychology, LMU Munich.

2.1.3 Stimuli and Apparatus

Stimuli were presented on one 17-inch monitor (Dell E178FPC, 75 Hz refresh rate) and on two LCD displays (84 × 48 pixels; diameter: 45 × 45 mm; Nokia 5110; with a PCD8544 controller by Philips), see Fig. 1. Throughout the entire experiment a picture of a static human face with a fixation cross between its eyes was displayed on the monitor. The LCD displays were attached to the monitor, one on each side of the monitor (15 cm from the middle of the screen, which amounted to 7.1° of visual angle). The target stimuli, either the capital letter F or T in black color, were presented on the LCD displays, covering 0.71° in height and 0.62° in width of visual angle. The participants were seated 120 cm from the monitor and a chin rest ensured that participants were sitting centered towards the monitor. Two approx. 15 cm long robot arms were fixed below the monitor and could point via two servo motors (continuous rotation servo motor, Par-

Fig. 1 Experimental setup of Experiment 1. In this example, the *left* robot arm is pointing to the left LED display. The static face in the *middle* was always the same and presented throughout all trials. The LED displays on which target letters were flashed are visible as little rectangles on the *left* and *right* side of the computer screen

allax Inc.) towards the LCD displays. A breadboard with 2 response keys (one marked as F and one as T) and a tone burst generator was positioned on a table in front of the participants.

In a separate room, a breadboard with a glowing LED and a response key that was also connected to the experimental setup was positioned in front of an experimenter. The setup was controlled by a microcontroller board (Arduino Mega 2560) with Hex-Schmitt-trigger-inverter (SN7414N, Texas Instrument), Hex-non-inverting buffer (HEF4050BP, NXP semiconductors) and voltage regulator (LM7805CT, Fairchild semiconductors).

2.1.4 Procedure

Before the experiment, participants were shown the setup, including the experimenter's room. Participants were briefed that they would engage in four blocks of the experiment and that after each block they had to distinguish whether the experiment was pre-programmed or human-controlled by an experimenter. Each block consisted of 80 trials, two of the blocks were human-controlled and two were pre-programmed. The order of human-controlled versus pre-programmed condition was randomized across participants. During the experiment participants were seated in front of
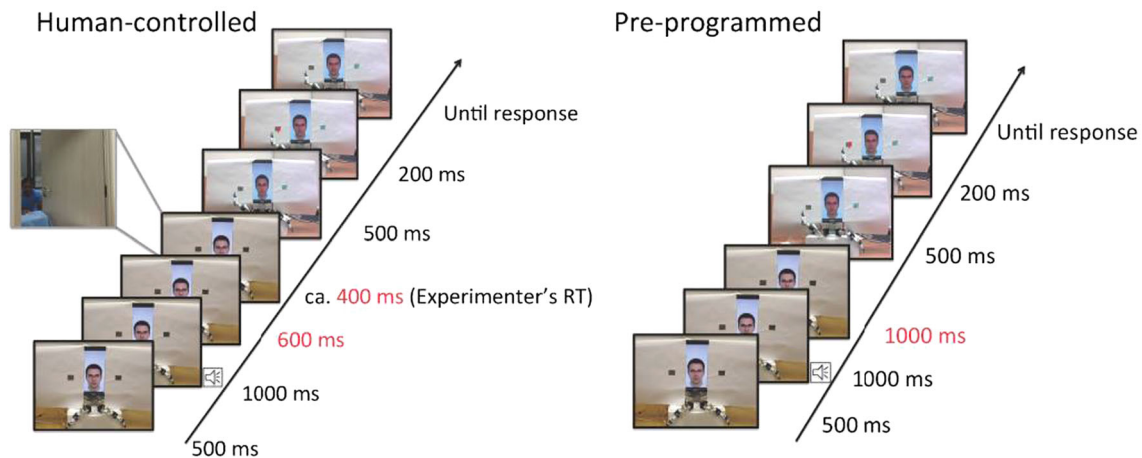
Fig. 2 Example trial sequence. Human-controlled condition (*left*) and the pre-programmed condition (*right*). A trial always started with a preparatory event lasting for 500 ms with robot arms in a "resting" position, face presented centrally and fixation cross between the eyes of the face. Subsequently, a beep was played for 1000 ms, signaling beginning of the trial. In the human-controlled condition, subsequent to the beep, a 600 ms break was introduced and then an LED lit up in the experimenter's room. The experimenter was asked to press the control key as fast as possible upon detecting the LED signal. In the pre-programmed condition (*right*), the break between the offset of the beep and the onset of the movement lasted for 1000 ms. Immediately after the experimenter's response (or after the 1000 ms break), a robot arm started moving either towards the *left* or towards the *right* LED display. The arm movement lasted for 500 ms until the arm reached its final position. After the arm movement ended, an F or T letter was flashed on one of the LED displays (either the one that the arm pointed to—valid trials; or the opposite display—invalid trials). The letter was present on the display for 200 ms. Participants responded to letter identity, and upon their response, the robot arm returned to the initial position and a trial ended

a computer monitor, wearing headphones to filter out background noise. The beep of the burst generator in front of them, however, was audible through the headphones. To start the experiment participants were instructed to press any key as soon as they were ready. Figure 2 presents the trial sequence. The trial started with a beep of 1000 ms. In the pre-programmed condition a pause of 1000 ms followed, after which either the right robot arm moved to point to the right LCD screen or the left robot arm moved to point to the left LCD screen (the right vs. left arm moved pseudo-randomly throughout a block of trials). After 500 ms the robot arm reached its final position and either the letter F or T flashed on one of the LCD displays for 200 ms (letter identity and side of presentation were pseudo-randomized throughout a block of trials). Participants were instructed to respond to the target stimulus as fast as possible by pressing left key for the letter "F", and right key for the letter "T". Reaction times (RTs) of participants were calculated from onset of target's appearance to onset of participants' response. As soon as the response was given, the robot arm moved down towards its starting position and the next trial started. Participants were told to keep their eyes fixated on the fixation cross displayed on the monitor throughout the whole trial and not to look at the LCD screens. They were instructed that they would have a high likelihood of missing the appearance of the target, if they moved their eyes to one of the LCD screens. With this instruction and manipulation, we aimed at measuring effects of covert attention, and not effects related to eye movements. The human-controlled condition differed the pre-programmed condition in that 600 ms after the beep, an LED in the experimenter's room glowed, signaling the experimenter to press the key which triggered an arm movement of the robot. Hence, the experimenter's task was to perform a simple reaction time task, by pressing the response key as fast as possible upon detecting the LED flash. As the whole room of the experimenter was occluded from the view of the participants, participants did not see the LED flashing. As soon as the experimenter pressed the response key, the robot arm moved. It was programmed such that the right or the left arm would move pseudo-randomly upon the response of the experimenter, similarly to the pre-programmed condition, in which the left or right arm also moved pseudo-randomly with equal likelihood for either of the two moving on a given trial. Mean RTs of the experimenter were estimated as amounting to ca. 400 ms, based on a pilot experiment. Therefore, in the human-controlled condition, the overall delay between the beep and the onset of the robot arm movement amounted to ca. 1000 ms on average, and thus was comparable to the pre-programmed condition. After the 1000 ms delay (relative to the beep) and the arm movement onset, the human-controlled condition continued in the same fashion as in the pre-programmed condition, see Fig. 2. RTs of the experimenter were recorded from onset of the LED glowing to onset of the experimenter's response.

After each block, participants had to fill out a brief questionnaire, in which they had to indicate if the block had been human-controlled or pre-programmed. As this experiment aimed at examining cognitive mechanisms involved in social interactions, an autism quotient (AQ) questionnaire [22]—which measures general social aptitude and autistic traits not only in clinical but also healthy populations—was administered in order to measure participants' social aptitude. Data of one participant were excluded from analyses due to an unusually high AQ score (AQ = 34; mean AQ score of remaining participants amounted to 20.05, SD = 5.6).

2.1.5 Data Analysis

2.1.5.1. Sensitivity to Human Behavior Sensitivity to human behavior was tested by comparing the accuracy of responses in the humanness judgment against chance, which in this case, i.e., two-alternative forced-choice task, was at the level of 50 %. Subsequently, we performed an analysis on the experimenter's mean, median RTs and standard deviations for each separate block in order to examine whether any of these would be predictive of participants' accuracy with respect to detecting the human behavior. We compared experimenter's mean, median RTs and SDs for blocks in which participants correctly detected human behavior vs. blocks in which participants responded erroneously claiming that the behavior was pre-programmed. For the analyses related to sensitivity to humanness, data of one participant were excluded due to presence of extremely long RTs of the experimenter (Max$_{RT}$ = 16 s) and thus very large standard deviation (M$_{SD}$ = 1999). Note that for the comparison between human-controlled blocks in which participants responded correctly versus erroneously, we could analyze data only of those participants who had responses in both types of blocks. That is, data of participants who were 100 % correct or 0 % correct could not be analyzed. Hence, data of only 9 participants entered these analyses.

2.1.5.2. Target Discrimination Performance First, the RT data were tested for normality through visual inspection of frequency distributions as well as through Kolmogorov–Smirnov (K–S) tests. The raw RT data were not normally distributed, as indicated by the positive skewness (67) of the distribution (Fig. 3) as well as a significant difference from the normal distribution, according to the K–S test: D (5245) = .451, p < .001. After exclusion of outliers (RTs shorter than 200 ms and longer than 1200 ms) the data remained not normally distributed (Fig. 4), D (3954) = .122, p < .001.[1]

---

[1] Note that after additional exclusion of RTs +/− 2 SD from means of individual participants or log-transformation of the data (after exclusion of RTs > 200 and < 1200), the RT data remained not normally distributed, D (3912) = .089, p < .001, and D (3912) = .027, p < .001, respectively.
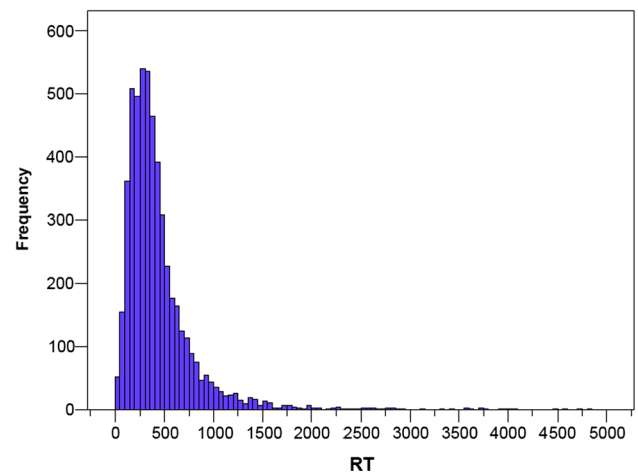

Fig. 3 Frequency distribution of participant RTs, Experiment 1
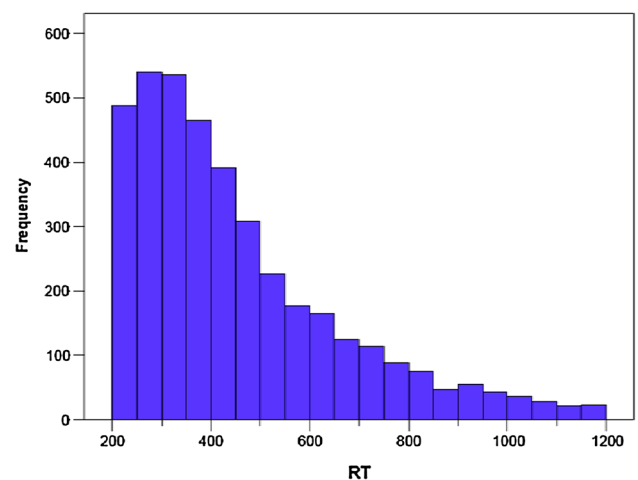

Fig. 4 Frequency distribution of participant RTs in Experiment 1 after exclusion of RTs < 1200 and >200 ms

The distribution of experimenter's RTs was also analyzed with respect to normality. Also these data were not distributed normally, D (2840) = .278, p < .001, Fig. 5.

As the RT data were not distributed normally (were positively skewed), we calculated median RTs for each participant, and each validity condition. Medians represent the central tendency better than means for skewed distributions. Before we calculated the median RTs, we excluded RTs < 200 ms, as they constituted erroneous key presses rather than actual reaction times. Two separate analyses were performed on the median RTs: one for the *actual* humanness (actual human-controlled vs. pre-programmed conditions, independent of participants' response in the humanness judgment) and the other for the *perceived* humanness (perceived human-controlled vs. preprogrammed, independent of actually presented condition) with validity (valid vs. invalid) and humanness (human-controlled vs. pre-programmed) as within-participants factors. Analogous analyses were per-
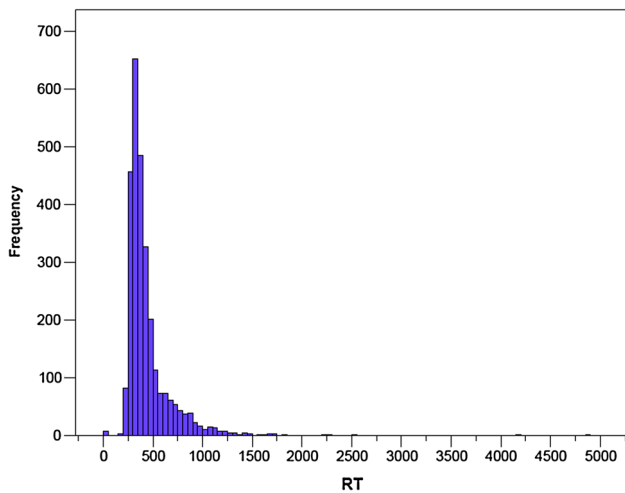
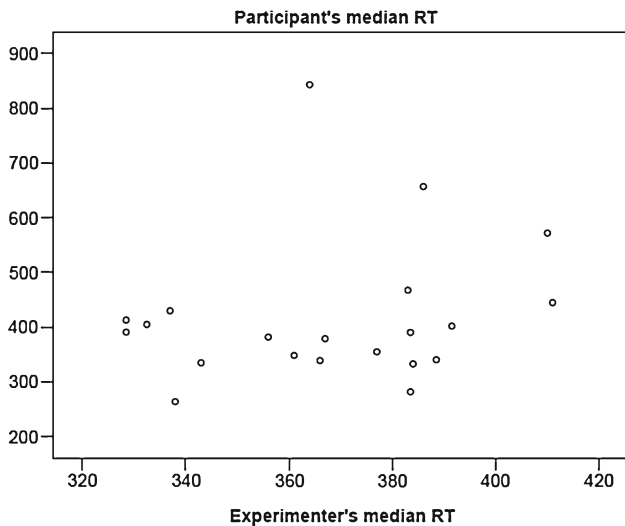**Fig. 5** Frequency distribution of experimenter's RTs in Experiment 1

Fig. 6 Scatter plot of the relationship between participants' median RTs and experimenter's median RTs. In line with the result of Pearson's correlation analysis, there is no trend for a linear regression between the scores

formed for error rates in the target discrimination task. To test the relationship between experimenter's RTs and participants' RTs, we correlated median experimenter's RTs with median participant RTs for each individual participant (Fig. 6).

2.2 Results

2.2.1 Sensitivity to Human's Behavior

Participants were able to detect human behavior with the average accuracy level of 64 %. This level of performance was significantly above chance, $t(19) = 2.85$, $p = .01$, two-

Table 1 Average median RTs (ms) as a function of validity and actual humanness together with the mean differences (M_{Diff}) between the validity conditions, and standard errors of the mean differences (SEM_{Diff})

	Human	Pre-programmed
Valid	425	383
Invalid	462	432
M_{Diff}	37	49
SEM_{Diff}	20	26

tailed. Neither the experimenter's mean RTs nor the median RTs did predict participants' response ("Mean/Median Perceived as human" = 441 ms/356 wms vs. "Mean/Median Perceived as pre-programmed" = 442/365 ms, p > .94 for the mean experimenter's RTs and p > .52 for the median experimenter's RTs). Interestingly, the standard deviations in experimenter's responses for an entire block were numerically larger ($M_{SD} = 267$ ms) when participants responded "human-controlled" than when they responded "pre-programmed" ($M_{SD} = 210$ ms), and this difference was marginally significant, $t(8) = 1.423$, $p = .096$, one tailed, $M_{Diff} = 57$ ms, $SEM_{Diff} = 40$.

2.2.2 Performance in Target Discrimination Task

2.2.2.1. RTs in Actual Humanness Conditions A 2 × 2 ANOVA with the factors validity (valid vs. invalid) and actual humanness (human-controlled vs. pre-programmed) on median RTs in the target discrimination task revealed the main effect of validity ($RT_{valid} = 404$ ms, SEM = 26 vs. $RT_{invalid} = 447$ ms, SEM = 36, $M_{Diff} = 43$ ms, $SEM_{Diff} = 18.83$), $F(1, 20) = 5.215$, $p = .033$, $\eta_p^2 = .207$. The effect of actual humanness ($RT_{human-controlled} = 443$ ms, SEM = 32 vs. $RT_{pre-programmed} = 407$ ms, SEM = 32, $M_{Diff} = 36$ ms, $SEM_{Diff} = 21$), was not significant, $p = .107$. The interaction between these two factors also did not reach the level of significance, $p = .693$, see Table 1.

2.2.2.2. RTs in Perceived Humanness Blocks A separate ANOVA with the factors validity (valid vs. invalid) and *perceived* humanness (human-controlled vs. pre-programmed) showed no significant effects or interactions, all $Fs < 3$, $ps > .1$. However, the median RTs in the blocks in which behavior was perceived as human were numerically faster on average (408 ms), relative to the blocks in which the behaviour was perceived as pre-programmed (431 ms), $M_{Diff} = 32$ ms, $SEM_{Diff} = 18$.

In the analysis of trials in which experimenter's RTs were > 350 and < 80 ms (mean RT of the experimenter = 401.09 ms, and not significantly different from the pre-programmed condition), none of the effects were significant,

all Fs < 1, ps > .35. Numerically, however, the perceived human-controlled blocks yielded faster median RTs (M = 409, SEM = 39) than the perceived pre-programmed blocks (M = 438, SEM = 26, M_{Diff} = 29 ms, SEM_{Diff} = 47, N = 9), which is in line with the pattern of results when all the trials were analyzed. Thus, analyzing only those trials, in which the experimenter's RTs were on average similar to the timing in the pre-programmed condition did not yield different results than analyzing all trials—a result similar to the "actual humanness" condition.

2.2.2.3. Relationship Between Experimenter's RTs and Participants' RTs. To test if the experimenter's RTs did not influence the participants' RTs, we conducted an analysis of correlation between the median RTs of the experimenter and median RTs of the participant for each individual participant. Pearson's correlation coefficient revealed that the experimenter's RTs were not significantly correlated with participants' RTs, r (19) = .204, p = .375, cf. Fig. 6.

2.2.2.4. Error Rates An analysis on error rates in the actual humanness condition revealed no significant effects or interactions, all $Fs < 3.5$, $ps > .08$. Numerically, the pattern of error rates paralleled RT results: less errors were committed in the valid trials (M = 5.7 %, SEM = 1.1) as compared to invalid trials (M = 8.6 %, SEM = 1.9, M_{Diff} = 2.9 %, SEM_{Diff} = 2.22); and less errors were committed in the pre-programmed (M = 5.7 %, SEM = 1.6) as compared to human-controlled condition (M = 8.6 %, SEM = 1.6, M_{Diff} = 2.9 %, SEM_{Diff} = 2.23). Analogous analysis on error rates in the perceived humanness blocks revealed no significant effects or interactions, all $Fs < 3$, $ps > .1$. Numerically, the pattern of error rates paralleled RT results: less errors were committed in the valid trials (M = 4.5 %, SEM = .9) as compared to invalid trials (M = 9.4 %, SEM = 2.7, M_{Diff} = 4.9 %, SEM_{Diff} = 2.84); and less errors were committed in the pre-programmed (M = 7.2 %, SEM = 1.2) as compared to human-controlled condition (M = 7.9 %, SEM = 1.9, M_{Diff} = 0.7 %, SEM_{Diff} = 0.66).

3 Experiment 2

Experiment 2 was designed in order to control for the appearance of the social agent. By substituting the static human face with a robot, we sought to test whether there might be any biases in Experiment 1 caused by the human face. Furthermore, in Experiment 2, the experimenter did not control the onset of the arm movements of the robot online, as in Experiment 1, but was only believed to do so (through instruction manipulation). In reality, the "human-controlled" condition was implemented through pre-recorded reaction times of an

experimenter, in order to have same onset variability across all participants.

3.1 Methods

3.1.1 Participants

Eighteen adult participants (Mean age: 24.6, SD: 3.47; six men) took part in the second experiment. All participants were healthy volunteers and had normal or corrected-to-normal vision. The experiment was conducted with the full understanding and written consent of each participant.

3.1.2 Ethics Statement

Experiment 2 was conducted at the Institute of Cognitive Systems, Technical University of Munich. All participants were healthy and adult. The experimental procedures consisted of purely behavioral data collection (RTs and error rates), and filling out two questionnaires. The procedures did not include invasive or potentially dangerous methods and were in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Data were stored and analyzed anonymously. Participants gave written consent and received monetary compensation or course credits for participating. These standard procedures are approved by the ethics committee of the Department of Psychology, LMU Munich.

3.1.3 Stimuli and Apparatus

Stimuli were presented on a 23-inch LCD monitor (iiyama ProLite B2409HDS, 60 Hz refresh rate). The target stimuli, either the capital letter F or T in white color, were presented on the monitor, covering 0.71° in height and 0.62° in width of visual angle. The participants were seated 125 cm from the monitor, and a chin rest ensured that participants were sitting centered towards the monitor. The participants responded by using a keyboard in front of them. A NAO robot (Aldebaran Robotics) was positioned in a sitting pose 20 cm in front of the screen with the head being positioned exactly in the middle of the screen. The robot's arms were used to point to the screen for the cuing paradigm, Fig. 7. In the resting position, both arms were raised up (as in Fig. 7, left arm). The cueing was performed by one of the arms bending backwards towards the respective part of the screen (Fig. 7, right arm). The target letter would appear between the location pointed by the arm and the head (so that the location of the target (left/right) would be cued by the arm, but the arm would not occlude the target), see Fig. 7. The robot's eyes were made to blink to signal the beginning of each trial. The experimenter sat in the same room as participants, but was occluded from the participants' view.

background noise. The trial started with the robot blinking its eyes for 300 ms. In the "pre-programmed" condition, a pause of 1065 ms followed, after which one of the arms of the robot (pseudo-randomly selected before the beginning of the experiment) would move to point to the screen behind him. In the human condition, this pause varied as explained above, with a mean of 1065 ms. 1000 ms after onset, the movement was finished, and either an F or a T (pseudo-randomly selected) would flash for 200 ms on the monitor. Participants were instructed to respond to this as fast and accurately as possible by either pressing the left key for F, or the right key for T. Their reaction times were measured from the onset of letter appearance to the pressing of the key. After the response, the robot moved its arm back into position, and the next trial started. As in Experiment 1, participants were asked to keep their eyes fixated on the face of the robot for the entire trial. After 24 trials, they were asked whether they believed the robot had been controlled by the experimenter or had been pre-programmed, and they answered by pressing a key (P for pre-programmed; H for human-controlled) on the keyboard. The keys were covered with colored stickers (with H/P written on them) for clarity. At the end of 21 blocks, participants filled a questionnaire where they indicated what cues led them to their choices. Additionally, an autism questionnaire [22] was administered, similarly to Experiment 1 (mean AQ score of participants: 15.83, SD = 5.59).[2]

3.1.5 Data Analysis

The first block for all participants served as practice, and was thus excluded from all analyses. Data of one participant were excluded from all analyses due to abnormally high reaction times in the target discrimination task (median RT = 1122, mean of other participants: 586 ms, SD = 116 ms). Data of an additional participant were excluded due to very high error rates in the letter discrimination task (10.8 % for the

Fig. 7 Experimental setup of Experiment 2. In this example, the robot is pointing to the *right* part of the screen with the *right* arm (*left* arm from the robot's perspective) by bending it backwards. The *left* arm (*right* arm from the robot's perspective) is in its "resting" position. This is an example of a valid trial, in which the target letter (F) appears also on the *right* side (same side as NAO's arm is pointing to)

3.1.4 Procedure

Before the experiment, participants were shown the setup and were briefed on both tasks of the experiment. They were told they would run 21 blocks where the robot movements might either be pre-programmed or online controlled by an experimenter. In reality, all movements were pre-programmed. Blocks labelled as "human-controlled" differed only in the variability of the timing of the movement onset of the robot's arms. For these blocks, we used a list of reaction times recorded before the experiment (mean RT = 465 ms, median RT = 438 ms, SD = 140 ms), while "pre-programmed" blocks used a single value (465 ms), corresponding to the mean of the human reaction times. As in Experiment 1, we added a fixed pause of 600 ms to these times before movement onset. During the experiment, participants were seated in front of the robot and the monitor, wearing headphones to filter out

---

[2] On a subset of eleven participants, a control condition was conducted after the end of the experiment (2 × 24 trials), to control for the overall performance in the letter discrimination task. In this condition, the robot did not move its arms and always remained in a neutral position after its eyes blinked to signal the start of a trial. There was also no question about the perceived humanness of the block; participants were instructed to only respond to the appearance of the letters in the screen. The same participant that was excluded from other analyses was also excluded from the analysis of the control condition data. Results showed that—when compared to the experimental conditions on the same subset of participants—the level of accuracy was similar and the RTs were slightly higher ($M_{RT} = 763$ ms, $M_{Accuracy} = 97\%$) versus "pre-programmed" condition ($M_{RT} = 685$ ms $M_{Accuracy} = 97$ %) or the "human-controlled" condition ($M_{RT} = 674$ ms, $M_{Accuracy} = 97$ %), but not significantly different, $F < .2$, $p > .8$, meaning that the cueing procedure did not impair overall performance in the target discrimination task.
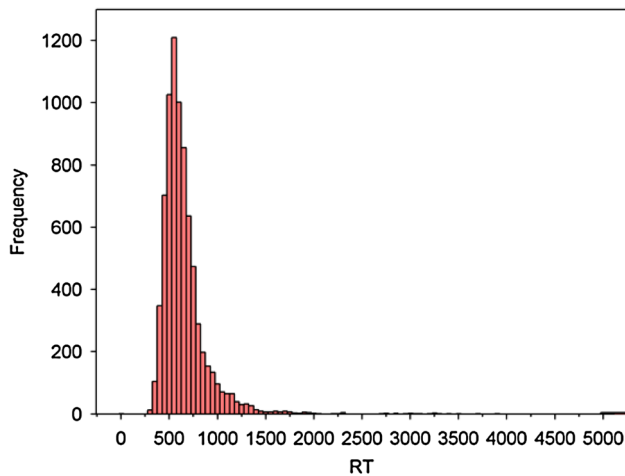
Fig. 8 Frequency distribution of participant RTs, Experiment 2

Fig. 9 Frequency distribution of participant RTs in Experiment 2, after exclusion of outliers below 200 ms and above 1200 ms

Fig. 10 Frequency distribution of experimenter's RTs

excluded subject, mean of other participants: 2.5 %, SD = 1.8 %).

3.1.5.1. Sensitivity to Human Behavior As in Experiment 1, we compared judgments of humanness against chance, and analyzed whether the mean/median reaction times or standard deviation in the "human-controlled" condition were predictive of participants' accuracy, by comparing blocks correctly identified as human to those falsely identified as "pre-programmed".

3.1.5.2. Target Discrimination Performance As in Experiment 1, RT data were tested for normality. Both subjects' and the experimenter's RT distributed were found to be significantly different from a normal distribution and had positive skewness. For the experimenter's RT: D (240) = .107, p = .008. For the raw participants' RT data: D (7860) = .177, p < .001. Outliers (RTs < 200 ms and > 1200 ms) made up 2.8 % of data. After excluding them, the RT data remained significantly different from a normal distribution: D (7464) = .090, p < .001. Figures 8 and 9 show the aggregated reaction time distributions of subjects before and after exclusion of outliers, respectively. Figure 10 shows the distribution of experimenter reaction times, as presented to the participants.

Median RTs as well as error rates were calculated for each participant and each validity condition, as in Experiment 1. Separate analyses were conducted for perceived humanness ("human-controlled" vs. "pre-programmed", as selected by the participants, independent of actual humanness condition) and for actual humanness ("human-controlled" vs. "pre-programmed", as presented to the participants, independent of their responses). We excluded trials where participants made a discrimination error, as well as those trials where RTs were below 200 ms or above 1200 ms (Fig. 9).
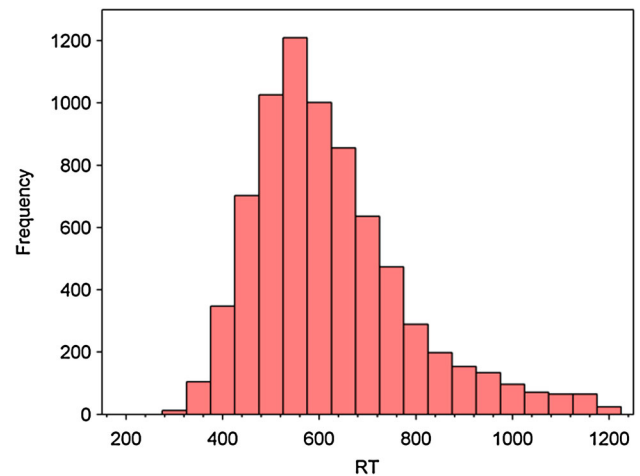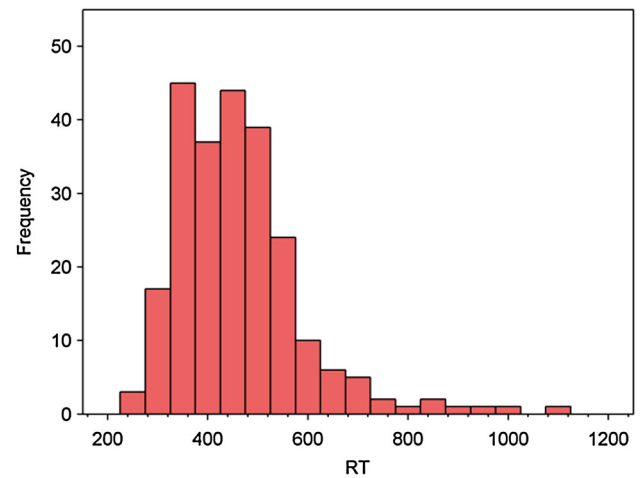
3.2 Results

3.2.1 Sensitivity to Human's Behavior

In Experiment 2, participants were also able to detect human behavior above chance, the average accuracy level being 57 % (SEM = 2.7). This level of performance was significantly above 50 % (chance level), $t(15) = 2.55$, $p = .022$, two-tailed. The accuracy in the humanness judgment (identifying the "human-controlled" condition correctly vs. perceiving it incorrectly as "pre-programmed") could be significantly predicted by the mean reaction time of the experimenter in the "human-controlled" blocks: (mean RT for error blocks: 454 ms, mean RT for correct blocks: 476 ms, $t(15) = 3.062$, p = .008). It was also significantly predicted by the median RT (average median RT for error blocks: 434 ms, mean RT for correct blocks: 444 ms, $t(15) = 2.217$, p = .043). In addition, it was also marginally predicted by the

Table 2 Average median RTs (ms) as a function of validity and actual humanness together with the mean differences (M_{Diff}) between the validity conditions, and standard errors of the mean differences (SEM_{Diff}) in Experiment 2

	Human	Pre-programmed
Valid	589	611
Invalid	594	609
M_{Diff}	6	2
SEM_{Diff}	28	28

Table 3 Average median RTs (ms) as a function of validity and perceived humanness together with the mean differences (M_{Diff}) between the validity conditions, and standard errors of the mean differences (SEM_{Diff}) in Experiment 2

	Human	Pre-programmed
Valid	601	596
Invalid	598	602
M_{Diff}	5	4
SEM_{Diff}	29	29

Table 4 Error rates (%) as a function of validity and actual humanness together with the mean differences (M_{Diff}) between the validity conditions, and standard errors of the mean differences (SEM_{Diff})

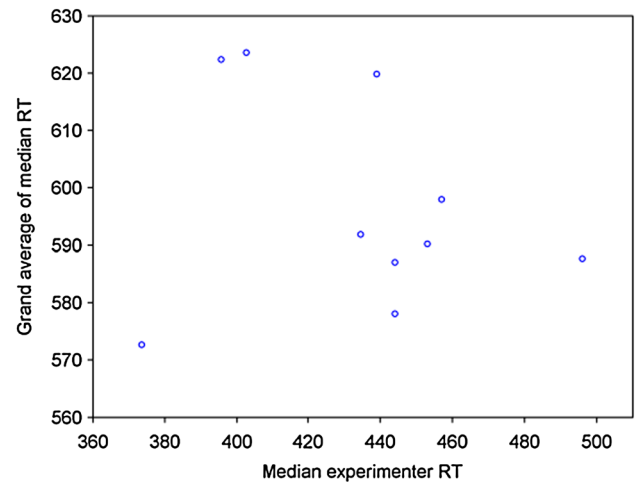|	Human	Pre-programmed
Valid	2.92	2.19
Invalid	2.19	2.81
M_{Diff}	0.73	0.62
SEM_{Diff}	0.7	0.7

Fig. 11 Average of median RTs across all subjects against experimenter's median RTs for each of the 10 human-controlled blocks. As in Experiment 1, there is no trend for a linear regression between the scores

standard deviation of the "human-controlled" blocks (mean of RT standard deviations for error blocks: 126 ms, for correct blocks: 147 ms, $t(15) = 1.877$, p = .04, one-tailed).

In summary, shorter reaction times of the experimenter and a lower variance during the block seems to have induced subjects to incorrectly categorize the block as pre-programmed.

3.2.2 Performance in Target Discrimination Task

3.2.2.1. RTs in Actual Humanness Conditions A 2 × 2 ANOVA with the factors validity (valid vs. invalid) and actual humanness on median RTs in the target discrimination task revealed a main effect of humanness, $F(1, 15) = 16.46$, $p = .001$, $\eta_p^2 = .523$ with short median RTs in the "human-controlled" condition ($M_{human-controlled} = 591$ ms, SEM = 14) relative to "pre-programmed" condition ($M_{pre-programmed} = 610$ ms, SEM = 14, $M_{Diff} = 18$ ms, $SEM_{Diff} = 20$). Other effects and interactions failed to reach the level of significance (all Fs < 1.9, all ps > .19).

3.2.2.2. RTs in Perceived Humanness Blocks A 2 × 2 ANOVA with the factors validity (valid vs. invalid) and *perceived* humanness showed no significant effects or interactions (all ps > .3).

Planned comparisons between the valid and invalid trials showed no significant differences for either the actual or the perceived humanness condition, ps > .148, one tailed (Tables 2, 3).

3.2.2.3. Error Rates Analyzing error rates in the target discrimination task revealed no significant effects in either perceived humanness (all Fs < .21, all ps > .65) or actual humanness (all Fs < 4.5, all ps > .05), cf. Table 4.

3.2.2.4 Relationship Between Experimenter's RTs and Participants' RTs As in Experiment 1, we correlated the median RTs of the experimenter and those of the participants. Because every participant was exposed to the same stimuli sequence, the analysis was conducted with the median RT for each block and the average of median RTs across all participants (Grand Average). No significant correlation was found, r (10) = −.205, p = .569, Fig. 11.

4 Discussion

Our study aimed at examining whether the human perceptual system is sensitive to subtle hints observable in behavior of others that can indicate that the behavior results from the operations of a human mind, rather than a non-intentional mechanistic device. This study was the first step to investigate the more general question of what types of information

humans use when they attribute mind and intentionality to an observed agent. In two experiments, we used a paradigm in which participants interacted with robots, which had two extensible arms. In Experiment 1, in some blocks the onset time of an arm movement was controlled by a computer program, in other blocks it was controlled by an experimenter, who was seated in a different room and was occluded from participants' view. In Experiment 2, the "human-controlled" condition was implemented through blocks in which the onset times of a movement were programmed but modeled after human reaction times, and participants were made to believe that an experimenter occluded from view was controlling the robot behavior is some blocks. In both experiments, participants were asked to perform a "Turing test", namely to determine whether they interacted with a human-controlled or a pre-programmed machine. Importantly, the only hint that participants could have possibly had regarding whether the behavior was human-controlled or pre-programmed was variability in onsets of the arm movements in the "human-controlled" condition (in the "pre-programmed" condition, the onset times were always fixed). The movement itself was, however, identical across conditions. Participants were not informed about what type of hints they should look for and base their judgment on. The crucial difference between Experiment 1 and Experiment 2 was that in Experiment 1, a human face was presented in the middle of the screen throughout the entire experimental procedure. In Experiment 2, participants observed the NAO robot performing the pointing gestures, and thus were not presented with human characteristics of appearance.

Results showed that in both experiments participants were able to detect the "human-controlled" condition with accuracy that was significantly above chance. This was independent of whether a human face or a robot face was presented in the middle of the setup, and thus the humanness judgment was not biased by characteristics of appearance of the stimuli. This suggests that the human perceptual system is sensitive to subtle characteristics of behavior (independent of appearance) that is typically human, i.e., behavior with a certain degree of variability. Interestingly, although in Experiment 1, mean reaction times of the experimenter (onset times of the pointing movements) did not predict whether the humanness judgment would be correct or not, in Experiment 2, the accuracy in humanness judgment did depend on mean/median onset times of the pointing movements and standard deviations. Interestingly, it was not the similarity of the mean/median onset times to the "pre-programmed" condition that made participants respond incorrectly as "preprogrammed" in the "human-controlled" condition (the correctly and erroneously judged blocks were equally similar to the preprogrammed condition in terms of mean/median onset times) but the actual onset delay. That is, participants were more likely to judge the "human-controlled" condition

as preprogrammed if the mean/median onset delay was short (Mean 454 ms; Median 434 ms), relative to when it was longer (Mean 476 ms; Median 444 ms). This shows that participants might have certain pre-assumptions (not necessarily at a conscious level) regarding what a "pre-programmed" condition would be like (faster), and this influenced their responses in the humanness judgment. Regarding the standard deviation within blocks (which was actually the hint that participants should have used for the humanness judgment), in Experiment 1, the standard deviations tended to predict participants' accuracy in the humanness judgments, while in Experiment 2, the standard deviations significantly predicted accuracy. More specifically, blocks in which the onset of the arm movement was more variable were more likely to be judged as "human-controlled" than blocks in which the variability was smaller. This suggests that participants might have made their humanness judgments based on how variable the movement onset was across trials within a block.

The finding that humans are able to detect subtle characteristics of other human behavior provides striking evidence that the human brain has certain assumptions and expectations concerning what a human-like behavior typically looks like, and can use these assumptions to discriminate human agents from non-agents. Similar results have been previously obtained with the use of dynamic stimuli that displayed biological motion [23–25], see [26] for a different account. In these studies, however, whole movement patterns and kinematics are typically available for participants' judgments. In contrast, our study is the first to show that very subtle hints and extremely reduced information in the pattern of behavior is enough to make participants discriminate human agents from non-agents.

Apart from the non-verbal Turing test, participants were also asked to discriminate target letters, which could either be validly or invalidly cued by the pointing movement of the robot. This manipulation was done in order to examine whether participants would be more ready to engage in joint attention with the robot when they perceived it as human-controlled (i.e., when they adopted the Intentional Stance), relative to when they perceived it as pre-programmed. This reasoning was based on previous studies [6,7] in which it was found that humans attend to where others gaze to a larger extent when they believe that the observed gaze behavior is controlled by a human (Intentional Stance adopted), relative to when they believe it is pre-programmed. The idea in the present paradigm was that if participants engage in joint attention with the robot, they should attend to where it points, and therefore discriminate the target letter better in valid trials, relative to invalid trials.

This was indeed what we observed in Experiment 1. Participants discriminated the target letter better (with faster RTs) when the robot pointed in the same direction (valid

trials), relative to when it pointed in the opposite direction (invalid trials). Interestingly, however, in Experiment 1, this validity effect was not modulated by whether the pointing movements were controlled by a human or by a computer program. Moreover, *perceived* humanness also did not affect the validity effects. Hence, it suggests that in Experiment 1, participants were equally likely to engage in joint attention with the robot, independent of whether its behavior was resulting from the operations of the human mind, or just a computer program.

This however, might have been due to the human face that was presented centrally throughout the experimental procedure. This might have biased participants into attending to where the robot arms pointed. This interpretation was partially confirmed by Experiment 2 in which a robot was observed and we did not find a significant effect of validity either in the actual or perceived humanness analyses. Therefore, the readiness to engage in joint attention might have been—to some extent—biased by whether participants observed stimuli with human appearance or not, which is in contrast to previous findings [11,27] in which validity effects were observed for both human and robot stimuli. The discrepancy between the present results in those reported in [11] might be due to the fact that in the present study, participants were exposed to an actual embodied robotic system, while in [11], participants observed only robot/human face stimuli presented on the screen. According to [28], an essential aspect of social cognition is real-time interaction, and therefore stimuli presented on a computer screen might not capture all aspects of social cognitive mechanisms. On the other hand, in [27] also an embodied robot was used and validity effects were observed for gaze-guided attentional orienting. Therefore, it might be that gaze is a stronger attention-guiding social hint than pointing gestures. Alternatively, the robot used in [27] was closer in appearance to a human.

In sum, there might be two different mechanisms influencing the process of attentional orienting in response to social directional cues exhibited by a robot. On the one hand, physical appearance might play some role in whether participants attend to where the robot points. On the other hand, the type of social cue might play a role, with gaze being a stronger cue than pointing. Future research needs to address these factors in more detail in a systematic manner.

In addition to the main effects of interest, we also observed that in Experiment 1, participants responded overall faster in the pre-programmed condition than in the human-controlled condition. This might have happened because the onset of movement was always at a fixed interval in the pre-programmed condition (relative to the offset of the preparatory beep), while this interval varied with the RTs of the experimenter in the human-controlled condition. This might have caused participants to have temporal expectations about when the target would appear, and hence the reaction times were overall faster in the condition in which they could have expected the target to appear. Importantly, however, longer RTs in the human-controlled condition (relative to pre-programmed) were presumably not due to that in some trials the experimenter might have responded very fast, and therefore might have not left enough time for participants to prepare for responding. This possibility has been excluded with analyzing data of those trials, in which RTs of the experimenter were on average not significantly different from the delay in the pre-programmed condition. In these trials, the pattern of results remained similar to the analysis of all trials. Interestingly, in Experiment 2, the main effect of humanness in the actual humanness analysis showed that participants' RTs were shorter in the "human-controlled" condition, relative to the "pre-programmed" condition (an opposite pattern of results than in Experiment 1). This therefore does not support the idea that variability in movement onsets was in general detrimental to target performance due to lack of precise temporal expectations. It might be that in Experiment 2, the "human-controlled" condition attracted attention to a higher degree or was in general more alerting than the "pre-programmed" condition, and hence shorter RTs. Therefore, multiple factors might have played a role in the general differences between RTs to target presentation in the "human-controlled" versus "pre-programmed" conditions; but there was no evident influence of any of these factors on the validity effect.

To conclude, the present results indicate that humans are tuned to detecting humanness in others' behavior—even when the hints concerning the human-like behavior are extremely subtle. This shows that the human brain has developed mechanisms for distinguishing other humans among the abundance of various dynamical systems that can behave in a manner very similar to humans. This is presumably a socially and evolutionarily important skill, as it might have allowed humans to detect other conspecifics based on characteristics of their behavior and movements, even when the entity's appearance was not visible.

These findings are of significance not only for theoretical considerations but also in terms of application for the emerging domain of social robotics. In this field, researchers aim at designing robots that are to interact with humans in daily lives [29–33]. Therefore, robot designers aim at producing robots that would be treated as true social interaction partners and not only as simple automata. Hence, one of the crucial questions in social robotics is whether robots need to look and behave very similarly to humans in order to be treated as socially acceptable interaction partners. Most attempts in social robotics go in the direction of creating robots whose appearance is very human-like [33–35]. However, Masahiro Mori has postulated the so-called uncanny valley hypothesis [36], according to which the rate of acceptance of robots

should be increasing with increasing resemblance of robots to humans up to a point, beyond which repulsive reactions might be observed, due to the uncanny feeling of "very similar to human" but still very different.

In any case, our findings show that in the attempts of making robots more and more similar to humans, social roboticists might need to focus more on subtleties of behavior rather than physical appearance of robots. This is in line with [37], where it has been shown that a combination of a human-like appearance with mechanical (non-biological) motion elicited specific response in a distinct brain region in the left posterior lateral temporal cortex, compared to conditions in which participants observed a human agent with biological motion or a robot with mechanical motion. The authors concluded that the human brain expects a human-like appearing entity to exhibit biological motion. Hence, movement characteristics of a robot are equally important as its physical appearance. Our findings extend the results reported in [37] by showing that introducing only simple variability to robot behavior might make the robots appear more human-like.

References

1. Baron-Cohen S (1995) Mindblindness: an essay on autism and theory of mind. MIT Press/ Bradford Books, Boston
2. Frith CD, Frith U (2006) How we predict what other people are going to do. Brain Res 1079:36–46
3. Dennett DC (2003) True believers: the intentional strategy and why it works. In: O'Connor T, Robb D (eds) Philosophy of mind: contemporary readings. Routledge, London, pp 370–390
4. Gallagher HL, Jack AI, Roepstorff A, Frith CD (2002) Imaging the intentional stance in a competitive game. Neuroimage 16:814–821
5. Chaminade T, Rosset D, Da Fonesca D, Nazarian B, Lutcher E, Cheng G, Deruelle C (2012) How do we think machines think? An fMRI study of alleged competition with an artificial intelligence. Front Hum Neurosci 6:1–9
6. Wiese E, Wykowska A, Zwickel J, Müller HJ (2012) I see what you mean: how attentional selection is shaped by ascribing intentions to others. PLoS One 7(9):e45391
7. Wykowska A, Wiese E, Prosser A, Müller H (2014) Beliefs about the minds of others influence how we process sensory information. PLoS One 9(4):e94339
8. Tomasello M (2010) Origins of human communication. MIT Press, Cambridge
9. Turing AM (1950) Computing machinery and intelligence. Mind 59:433–460
10. Pfeiffer UJ, Timmermans B, Bente G, Vogeley K, Schilbach L (2011) A non-verbal Turing test: differentiating mind from machine in gaze-based social interaction. PLoS One 6(11):e27591
11. Chaminade T, Okka MM (2013) Comparing the effect of humanoid and human face for the spatial orientation of attention. Front Neurorobot 7:12
12. Friesen CK, Kingstone A (1998) The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. Psychon Bull Rev 5:490–495
13. Driver J, Davis G, Ricciardelli P, Kidd P, Maxwell E, Baron-Cohen S (1999) Gaze perception triggers reflexive visuospatial orienting. Vis Cogn 6:509–540
14. Frischen A, Bayliss AP, Tipper SP (2007) Gaze cuing of attention: visual attention, social cognition, and individual differences. Psychol Bull 133:694–724
15. Liebal K, Tomasello M (2009) Infants appreciate the social intention behind a pointing gesture: commentary on "Children's understanding of communicative intentions in the middle of the second year of life" by Aureli T, Perucchini P, Genco, M. Cogn Dev 24:13–15
16. Liszkowski U, Carpenter M, Henning A, Striano T, Tomasello M (2004) 12-month-olds point to share attention and interest. Dev Sci 7:297–307
17. Liszkowski U, Carpenter A, Striano T, Tomasello M (2006) 12 and 18-month-olds point to provide information for others. J Cogn Dev 7:173–187
18. Grassmann S, Tomasello M (2010) Young children follow pointing over words in interpreting acts of reference. Dev Sci 13:252–263
19. Cabibihan JJ, So W-C, Saj S, Zhang Z (2012) Telerobotic pointing gestures shape human spatial cognition. Int J Soc Robot 4:263–272
20. Cabibihan JJ, So WC, Nazar MA, Ge SS (2009) Pointing gestures for robot-mediated communication interface. In: Xie M et al (eds) Intelligent robotics and applications, vol 5928. Springer, Berlin / Heidelberg, pp 67–77
21. Ham J, Bokhorst R, Cuijpers R, van der Pol D, Cabibihan JJ (2011) Making robots persuasive: the influence of combining persuasive strategies (gazing and gestures) by a storytelling robot on its persuasive power, social robotics. Springer, Berlin / Heidelberg
22. 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. J Autism Dev Disord 31:5–17
23. Johansson G (1973) Visual perception of biological motion and a model for its analysis. Percept Psychophys 14:201–211
24. Grossman E, Blake R (2002) Brain areas active during visual perception of biological motion. Neuron 35:1157–1165
25. Thornton IM, Vuong QC (2004) Incidental processing of biological motion. Curr Biol 14:1084–1089
26. Hiris E (2007) Detection of biological and nonbiological motion. J Vis 7:1–16
27. Wykowska A, Kajopoulos J, Ramirez-Amaro K, Cheng G. (accepted). Autistic traits and sensitivity to human-like features of robot behavior. Interact Stud
28. Schilbach L, Timmermans B, Reddy V, Costall A, Bente G, Schlicht T, Volgeley K (2013) Toward a second-person neuroscience. Behav Brain Sci 36:393–462
29. Li H, Cabibihan JJ, Tan YK (2011) Towards an effective design of social robots. Int J Soc Robot 3:333–335
30. Brayda L, Chellali R (2012) Measuring human–robots interactions. Int J Soc Robot 4:119–221 (Editorial)
31. Cabibihan JJ, Williams MA, Simmons R (2014) When robots engage humans. Int J Soc Robot 6:311–313. doi:10.1007/s12369-014-0249-8
32. Cabibihan JJ, Wing-Chee S, Pramanik S (2012) Human-recognizable robotic gestures. IEEE T Auton Mental Dev 4:305–314
33. Yagoda RE, Gillan DJ (2012) You want me to trust a ROBOT? The development of a human–robot interaction trust scale. Int J Soc Robot 4:235–248
34. Ishiguro H, Nishio S (2007) Building artificial humans to understand humans. J Artif Organs 10:133–142

35. Nishio S, Ishiguro H, Hagita N (2007) Geminoid: teleoperated android of an existing person. In: de Pina Filho AC et al (eds) Humanoid robots: new developments. I-Tech Education and Publishing, Vienna, pp 343–352
36. Mori M (1970) Bukimi no tani. The uncanny valley (MacDorman KF, Minato T, trans.). Energy 7:33–35
37. Saygin AP, Chaminade T, Ishiguro H, Driver J, Frith C (2012) The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. Soc Cogn Affect Neurosci 7:413–422

Agnieszka Wykowska is a senior researcher 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 her 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). She has been involved in a Cluster of Excellence Cognition Technical Systems (CoTeSys, 2006–2010)—an interdisciplinary initiative aiming at designing robots with cognitive capabilities. She is an Associate Editor of the International Journal of Social Robotics, and she has been a member of the Program Committee of the International Conference on Social Robotics (2010, 2012, 2013, 2014). Her research interests include visual attention and perception, action planning, social cognition and perception, human-robot interaction and social robotics. In her research, she uses psychophysics and EEG/ERP methodology.

Jasmin Kajopoulos is a M.Sc. student of the elite Neuro-cognitive Psychology (NCP) study program of the Ludwig-Maximilians-Universität München (LMU), Germany. She obtained her B.Sc. degree in biology from the LMU and did her bachelor's thesis in neurobiology. Since 2012 she has been involved in interdisciplinary projects, aiming at understanding human-robot interaction and social neuroscience through robotics. To that end, she has also been working at the Institute for Cognitive Systems, Technische Universität München. Currently, she works—under the supervision of Dr. Agnieszka Wykowska and Dr. Tan Yeow Kee—on a collaborative project between the LMU and the Agency for Science, Technology and Research (A*Star) in Singapore, where she aims at understanding social robotics and its use in therapy.

Miguel Obando-Leitón is a Ph.D. Student of the Graduate School of Systemic Neurosciences of the Ludwig-Maximilians-Universität (LMU) München, Germany under the supervision of Prof. Werner Hemmert. He is currently working in the Department for Bio-inspired Information Processing of the Technische Universität München on the development of better coding strategies for cochlear implants. His background is Physics (B.Sc. from the Karlsruher Institut für Technologie, Germany) and Neurosciences (M.Sc. from the LMU, Master thesis under supervision of Dr. Agnieszka Wykowska). His main interest areas are human sensory perception, computational neuroscience and its technical applications.

Sushil Singh Chauhan received his Master of Science degree in Electrical Engineering with specialization in Automation and Control from the National University of Singapore in 2013 and his Bachelor of Technology degree in Electronics and Instrumentation from Vellore Institute of Technology, India in 2011. He is currently employed as Research Assistant at SINAPSE (Singapore Institute for Neurotechnology) where he worked on peripheral nerve injury to build/program implantable devices like muscle stimulators for animal studies. He also worked as Research Engineer at NUS Department of Electrical and Computer Engineering. His research interests are in the field of biomedical engineering, physiological measurement, social robotics and robotics for autism assessment and early intervention.

John-John Cabibihan was conferred with a Ph.D. in the area of biomedical robotics by the Scuola Superiore Sant'Anna, Pisa, Italy in 2007. Concurrent with his Ph.D. studies, he was awarded an international scholarship grant in 2004 by the Ecole Normale Supérieure de Cachan, France. Therein, he spent one year at the Laboratoire de Mécanique et Technologie. From 2008 to 2013, he was an Assistant Professor at the Electrical and Computer Engineering Department of the National University of Singapore, where he also served as the Deputy Director of the Social Robotics Laboratory and an Affiliate Faculty Member at the Singapore Institute of Neurotechnologies (SiNAPSE). He is presently an Associate Professor at the Mechanical and Industrial Engineering Department of Qatar University. He serves at the Editorial Boards of the International Journal of Social Robotics, International Journal of Advanced Robotics Systems, Frontiers in Bionics and Biomimetics, and Computational Cognitive Science. He is a Senior Member of the IEEE and was a past Chair of the IEEE Systems, Man and Cybernetics Society, Singapore Chapter (terms: 2011 and 2012). He was also the General Chair of the 6th IEEE International Conference on Cybernetics and Intelligent Systems (2013), Program Chair of the 4th International Conference on Social Robotics (2012), and Program Co-Chair of the 2nd International Conference on Social Robotics (2010). Over the years, his work has been focused toward assistive and social robotics for the therapy of children with autism, lifelike prosthetics, bio-inspired tactile-sensing, and human-robotic touch and gestures.

Gordon Cheng is the Professor & Chair of Cognitive Systems, Founder and Director of the Institute for Cognitive Systems, Technische Universität München. Formerly, he was the Head of the Department of Humanoid Robotics and Computational Neuroscience (2002–2008), ATR Computational Neuroscience Laboratories, Kyoto, Japan. He was the Group Leader (2004–2008) for the JST International Cooperative Research Project (ICORP), Computational Brain. He has also been designated as a Project Leader (2007–2008) for National Institute of Information and Communications Technology (NICT) of Japan. He holds visiting professorships worldwide in multidisciplinary fields comprising: Mechatronics (France), NeuroEngineering (Brazil) and Computer Science (USA). He held fellowships from the Center of Excellence (COE), Science, and Technology Agency (STA) of Japan. These fellowships were taken at the Humanoid Interaction Laboratory, Intelligent Systems Division at the ElectroTechnical Laboratory (ETL), Japan. He received a PhD in Systems Engineering (2001) from the Department of Systems Engineering, Australian National University, and Bachelor (1991) and Master (1993) degrees in Computer Science from the University of Wollongong, Australia. He was also the Managing Director G.T.I. Computing in Australia. His research interests include humanoid robotics, cognitive systems, brain-machine interfaces, human-robot interaction and social robotics. Prof. Cheng is the co-inventor of approx. 15 patents and has co-authored approx. 250 technical publications, editorials and book chapters. He is a senior member of the IEEE Robotics and Automation and Computer Society. He is on the editorial board of the International Journal of Humanoid Robotics.