Int J of Soc Robotics
DOI 10.1007/s12369-014-0231-5

Do Robot Performance and Behavioral Style affect Human Trust?

A Multi-Method Approach

Rik van den Brule · Ron Dotsch · Gijsbert Bijlstra · Daniel H. J. Wigboldus · Pim Haselager

Accepted: 1 March 2014

Abstract An important aspect of a robot's social behavior is to convey the right amount of trustworthiness. Task performance has shown to be an important source for trustworthiness judgments. Here, we argue that factors such as a robot's behavioral style can play an important role as well. Our approach to studying the effects of a robot's performance and behavioral style on human trust involves experiments with simulated robots in video human–robot interaction (VHRI) and immersive virtual environments (IVE). Although VHRI and IVE settings cannot substitute for the genuine interaction with a real robot, they can provide useful complementary approaches to experimental research in social human robot interaction. VHRI enables rapid prototyping of robot behaviors. Simulating human–robot interaction in IVEs can be a useful tool for measuring human responses to robots and help avoid the many constraints caused by real-world hardware. However, there are also difficulties with the generalization of results from one setting (e.g., VHRI) to another (e.g. IVE or the real world), which we discuss. In this paper, we use animated robot avatars in VHRI to rapidly identify robot behavioral styles that affect human trust assessment of the robot. In a subsequent study, we use an IVE to measure behavioral interaction between humans and an animated robot avatar equipped with behaviors from the VHRI experiment. Our findings reconfirm that a robot's task performance influences its trustworthiness, but the effect of the behavioral style identified in the VHRI study did not influence the robot's trustworthiness in the IVE study.

Keywords Social robotics · Trust · Video stimuli · Immersive virtual environments

Electronic supplementary material The online version of this article (doi:10.1007/s12369-014-0231-5) contains supplementary material, which is available to authorized users.

R. van den Brule (✉) · P. Haselager
Donders Institute for Brain, Cognition and Behaviour,
Radboud University Nijmegen, Montessorilaan 3,
6525 HR Nijmegen, Netherlands
e-mail: r.vandenbrule@donders.ru.nl

R. van den Brule · R. Dotsch · G. Bijlstra · D. H. J. Wigboldus
Behavioural Science Institute, Radboud University Nijmegen,
Montessorilaan 3, 6525 HR Nijmegen, Netherlands

1 Introduction

Increasingly advanced robots are being developed to aid in households. The introduction of these autonomous domestic robots into our social environment poses novel challenges to robot designers. People readily attribute anthropomorphic qualities to domestic robots and treat them as social agents, even when these robots were not designed with this in mind [1]. Domestic robot designers might improve user experience and acceptance of their products by taking this human desire for social interaction into account.

An important aspect of social behavior is the ability to convey and perceive trustworthiness. Trust is a fundamental concept in sociology [2], neuroeconomics [3], management studies [4,5], and psychology [6], and is described as a dyadic (interpersonal) relationship between a party that has to be trusted (the trustee) by another (the trustor) [4,7]. Trust is also described as the belief that a trustee will assist a trustor in reaching a common goal [6].

In a domestic setting, humans will engage in trust relationships with robots that perform household chores in which the robots are the trustees, and in which their owners take the role of the trustor. Importantly, domestic robots will ostensibly be able to perform a variety of tasks, not all of which they will be able to perform at the same level of autonomy

Published online: 27 March 2014

and/or expertise. Robots may be fully proficient at some sets of tasks, but at others only sufficient, or sufficient under guidance, or simply incapable. For instance, robots may be very competent at mowing lawns, but less so at doing the dishes. Also, because households can be quite dynamic, a domestic robot might encounter situations during routine operations in which it can no longer ensure a good outcome of its actions. An important way to prevent misunderstandings or even accidents, is to equip domestic robots with the means to convey their to be expected level of trustworthiness on a task in a natural way.

Trustworthiness of a trustee can be construed using several sources. Task performance could be the most objective source for trust, because in general previous behavior is a good predictor of future outcomes [8]. Consequently, ideally trustworthiness is estimated from prior observation. However, task performance is difficult to determine without previous interaction or knowledge of the robot.

Aside from task performance, trust is based also on more immediately observable, yet more subjective factors. Trust arises not only from *how well* a task is performed, but also from *the way* it is performed. This more subjective factor can be split into appearance [9–11] and behavioral style.

Appearance is the way a trustee (e.g. a robot) looks, and cannot be easily changed. In contrast, we define behavioral style as all the externally observable nonverbal aspects of the way a robot behaves while engaged with a task, such as the trustee's body language, facial expression, looking behavior or the way the trustee moves and acts. Nonverbal behavioral cues of confidence or doubt can be easily displayed and changed, and this should affect the trustworthiness interpretation of a trustee. For instance, it has been shown that interviewer manners, such as position and gestures, have a significant effect on the interviewer's perceived trustworthiness [12,13].

A recent meta-analysis on human–robot trust relationships found that a robot's performance aspects, such as its reliability and predictability, were found to have the biggest influence on trust. However, other robot attributes, such as its level of anthropomorphism, and personality, were also found to play a small but sizeable role in the development of trust [14].

Unfortunately, to our knowledge, no research exists in an HRI context regarding the way in which trustworthiness information is derived from nonverbal behavioral styles of trustees during social interaction (but see [15]). The main goal in this research therefore is to identify behavioral styles that robots are able to display so that they are noticed by humans, and that may affect the extent to which a robot is trusted.

Subjective sources of trustworthiness are especially useful when task performance information is unavailable. They can be used to develop an initial estimation of a trustee's trustworthiness prior to the initiation of interaction. According to dual process models used in social psychology [16,17], these subjective sources are processed automatically and unconsciously, while more explicit sources such as task performance are processed more deliberately and consciously.

Table 1 Calibration of trust

Perceived trustworthiness	Actual trustworthiness	
	Low	High
High	Overtrust (overreliance)	Calibrated High trust
Low	Calibrated Low trust	Undertrust (underreliance)

When the levels of task performance and behavioral cues are aligned, trust in the robot is calibrated. When the level of trustworthiness conveyed by the robot's appearance is lower than the trust generated by the robot's task performance, the robot will be undertrusted by a trustor. The opposite effect, when too much trust is placed in the robot based on its behavioral cues than its task performance warrants, the robot is being overtrusted

It has been shown that people tend to make riskier choices in economical exchange games such as the trust game [18][1] when their interaction partner has a trustworthy face compared to a partner with an untrustworthy face, despite the performance of both partners being the same [19]. This effect is still present after many iterations of the game, even as the trustworthiness estimation is updated to take the more predictive information of task performance into account [20].

In order to manage the expectations of human users, robots should convey (changes in) the extent to which they can be trusted to perform their task competently, i.e., their level of trustworthiness. A proper calibration of task performance and trustworthy behavioral style is favorable for high quality human–robot interaction. When a robot is or becomes bad at performing a certain task, it would be useful to convey a level of uncertainty or doubt to its owner so that trusting the robot too much (called overtrust) is avoided (see also Table 1). As a consequence, its owner may anticipate the robot's potential mistakes and thereby behave appropriately (e.g. preemptively). Similarly, if the robot anticipates performing a task well, it should be able to convey this to its user, thereby avoiding too little trust, (called undertrust), so that the user does not waste time and energy by monitoring the robot unnecessarily.

[1] In the trust game [18], participants must decide whether to invest money in a partner in an uncertain context. The partner receives this money, multiplied by a factor (usually 3 or 4). The partner must then decide whether to reciprocate the trust by sending back some of the money, resulting in a net gain for both players, or keeping all the money for themselves.

2 HRI Research Paradigms

Currently, two experimental paradigms are used in social HRI, namely experiments with real world robots (e.g., [21–23]), in which participants interact with a robot in real-life situations, and video experiments (also known as Video HRI or VHRI), in which participants watch filmed or animated, computer generated stimuli of robots (e.g., [24–28]). While both approaches have great value, they also have disadvantages. Real world robots are often too complex to allow a quick and cheap change of their appearance or behavior in detail. Experiments with video stimuli are becoming an increasingly common complementary method in HRI research [29], mainly to rapidly prototype and experiment with different robot designs. However, filmed stimuli do not permit the type of interactivity that is so characteristic for normal social interaction.

Immersive Virtual Environments (IVEs), also known as Virtual Reality, provides a research methodology that falls nicely in between video stimuli and real robot interaction. IVEs provide an immersive experience by immersing participants in a virtual, computer generated world. This is usually done by tracking participants' head movements and providing stereoscopic 3D computer graphics through a Head Mounted Display (HMD) (for a technology overview, see [30]). Any real or imaginary environment can be (re)created as an IVE, limited only by the bounds of imagination and computer memory. With the use of Immersive Virtual Environment Technology (IVET), participants can interact with computer-generated robots that may be too complex (too expensive and/or too time-consuming) to build or program in real-life.

IVEs are used as a methodological tool in social psychology to create experimental settings that are both controlled and realistic [30]. Moreover, it is easy to measure human behavior continuously and unobtrusively as IVEs keep track of a participant's location, orientation and pose. There are, of course, limitations to the realism of IVEs [31]. However, even though every participant is aware that the world they perceive is artificial, participants still exhibit the same automatic behavioral responses as they would in the real world [32–34]. This not only makes it possible to create a more dynamic interaction between humans and virtual robots, behavioral metrics can also be measured and used in data analysis.

Because of these reasons, IVET can benefit the field of social robotics as well. IVEs can be viewed as an extension of a robot simulator with real-time data of human behavior. Robot simulators are a cheaper and more time efficient to test algorithms and embodiments for robots [35]. Although video and IVET based experiments cannot substitute the full experience of interacting with a real robot and participants prefer to interact with a real robot rather than watching a video [36], video and IVE experiments can provide a useful complementary methodology for social HRI [24,28,36]. We feel that IVET fills a methodological gap between real world HRI and VHRI, because it allows for dynamic interaction without the need to build a real robot.

3 Overview

In this paper, we attempt to show how VHRI and IVE studies can be used as complementary research paradigms. We decided to make use of video stimuli to rapidly prototype different robot behavioral styles that subsequently were further studied in a more interactive IVE setting. In an exploratory VHRI study with computer generated video stimuli of a humanoid robot, we identified trustworthy and untrustworthy behavioral styles (Experiment 1). Next, we made use of IVET to validate behavioral styles identified in the first experiment in a setting where participants were also occupied with their own task, which offers a more realistic scenario for Human Robot Interaction in a domestic setting. This setup also enabled us to measure participants' monitoring of the robot, which potentially can be used as an unobtrusive measure of trust (Experiment 2).

Trustworthiness judgments of robots can be measured in several ways. One possibility is to use questionnaires to explicitly ask people how much they trust the robot [37]. It is also possible to derive trustworthiness judgments from the score a team obtains in a game in which trust is an important factor for success, as is done in many research concerning economic exchange [15,18]. Additionally, behavior of a person expressive of his trust in the robot can be recorded. Examples of such behavior are the amount of attention the robot requires [38] and proxemics [39].

4 General Method

4.1 Task Context: The Van Halen Task

Trust is critical in a collaborative task context. We therefore devised a collaborative game in which a trust relationship between players can emerge. A player of the task (e.g. the robot trustee) sits in front of a conveyor belt. Differently colored balls appear on the right side of the belt, and move towards the other side of the conveyor at a constant speed. The player's goal is to remove the brown colored balls from the conveyor, thereby gaining as high a score as possible. We named this task the Van Halen Task, after the famous rock band that demanded a bowl of M&Ms be placed back stage in their dressing room, but without the brown M&Ms, whenever they performed.

Two types of mistakes can be made: a brown ball might be missed and allowed to reach the end of the conveyer, or a non-brown ball might be incorrectly removed from the conveyor.

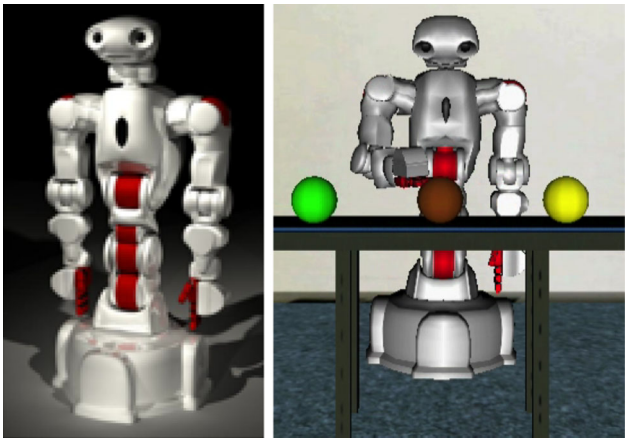

Fig. 1 The virtual robot used in the experiments was inspired by TWENDY-one [40]. *Left panel* a high end rendering of the virtual robot model. *Right panel* the virtual robot performing the Van Halen task

In the interactive version of the Van Halen Task (Experiment 2), a participant and the robot perform the task simultaneously. Trust is of importance in this task because the participants have the opportunity to correct the robot's mistakes, at the expense of their own performance on the task.

4.2 Virtual Robot

The appearance of the robot was based on TWENDY-one, a prototype of a robotic assistant for elderly people [40]. This anthropomorphic robot was designed with friendliness in mind. Some small changes were made to the 3D model of TWENDY-one. For instance, our robot's head included a mouth (see Fig. 1). In the IVE, the robot was slightly larger than its real-world counterpart so that participants could look the robot more or less straight in the eyes.

5 Experiment 1

5.1 Purpose

In Experiment 1, VHRI was used to rapidly prototype and identify which behavioral styles of a robot affect its trustworthiness in the context of the Van Halen task, as the interpretation of nonverbal behavioral cues are in general highly context dependent [41]. Additionally, we assessed whether different behavioral styles of the robot have different effects on trustworthiness judgments as a function of task performance. In addition to the robot's task performance, three behavioral aspects of the robot were manipulated in the experiment and were expected to affect trustworthiness judgments: the robot's gaze behavior, its motion fluency, and its hesitation when it reached for balls on the conveyor (see Sect. 5.2.2 for details). Gaze behavior is a key element of social interaction between agents [42], and is often used in HRI [43]. We chose motion fluency because agents that presents trembling motions would be considered less trustworthy; and hesitating behavior because an agent employing hesitating motions can convey the sense that the agent is uncertain about its upcoming action. The nature of this experiment was exploratory.

5.2 Method

5.2.1 Participants

160 participants were recruited at the Faculty of Social Sciences of the Radboud University Nijmegen. Participants were rewarded with a candy bar of their choice. Four participants were excluded before analyzing the data, three because they indicated they had prior knowledge about the experiment and one because of a computer error. This left 156 participants for analysis (35 male, 121 female),[2] median age: 21, age range: 18–52.

5.2.2 Materials and Design

Different video clips of the robot performing the Van Halen Task were created with WorldViz Vizard 3.0 3D rendering software. In each clip, 14 balls, of which four were brown, passed on a conveyor belt. Each clip was approximately 40 s in duration.

For each of the three robot behaviors two styles were created, which could be either enabled or disabled independently. For gaze behavior, we created a style in which the robot moved its head following the balls on the conveyor, and a style in which the robot looked at the conveyor without moving its head. For fluency of movement, we created a smooth style, in which the robot would fluently move its arm to reach for the balls on the conveyor, and a trembling style, in which the robot's arm would shake rapidly, resembling a trembling movement, whenever it reached for a ball. For hesitating behavior, we had the robot pull back its arm twice during the pickup movement of the balls before eventually picking it up. All behavioral styles were performed for the entire duration of the clip.

We included two levels of task performance. In the good performance condition, the robot made no mistakes. In the relatively bad performance condition, the robot made some mistakes: it missed the last two of the four four brown balls (which did not follow each other immediately) and instead picked up the ball before or after those.

[2] The gender ratio of the participant pools of both experiments is skewed towards female. Although this has no consequence when comparing the results from the two experiments in this work, it may limit the external validity of these studies.

Table 2 *t* test results of the manipulation checks

Manip. check	$t(154)$	$p <$	r	Manip. off		Manip. on	
				M	(SD)	M	(SD)
Looking	3.48	.001	.27	4.73	(1.66)	5.61	(1.47)
Fluency	17.87	.001	.82	2.65	(1.54)	6.22	(.86)
Hesitation	6.17	.001	.45	3.10	(1.73)	4.77	(1.64)
Performance	18.50	.001	.83	6.09	(1.18)	2.61	(1.17)

Note: That the manipulation check for task performance was reverse coded; i.e. a high score means participants noticed few mistakes

This setup resulted in a 2 (gaze behavior: look at conveyor vs. follow balls) × 2 (motion fluency: trembling vs. smooth) × 2 (hesitation: hesitate vs. no hesitate) × 2 (task performance: some mistakes vs. no mistakes) between subject design. So, in total, there were 16 conditions in the experiment.

5.2.3 Procedure

To establish a baseline context of robot behavior, all participants first watched a video of a robot performing the Van Halen task with its gaze looking at the conveyor, smooth motion fluency, no hesitating behavior, and good task performance. Participants then watched one of the 16 video segments in which we manipulated robot behavior and task performance (one of the clips was exactly the same as the baseline clip, to ensure a full design for later analyses). The robot in this second clip had a different color from the baseline, implying that it was a different robot than the one from the first video.

After the second video clip participants answered several questions about the second robot, using 7-point Likert scales, with 1 meaning "not at all", and 7 meaning "extremely". These questions included manipulation checks (e.g., 'How much did you find the robot to tremble?'; 'How much attention did the robot pay to the balls on the conveyor?'), and a question about the robot's trustworthiness ('How trustworthy did the robot appear to you'). We also included a calibration measure of how well the robot's behavior was aligned with its task performance by asking participant how surprised they were by the performance of the robot given their impression based on its behavior.[3]

5.3 Results

5.3.1 Manipulation Checks

Independent sample *t*-tests on the manipulation checks showed that participants noticed all manipulations, all *t*s > 3.40, all *p*s $< .001$. As can be seen in detail in Table 2, participants rated the robot lower on the control question for each manipulation when the behavior was turned off compared to when it was turned on.[4]

5.3.2 Trustworthiness

A 2 (gaze behavior: look down vs. follow balls) × 2 (motion fluency: smooth vs. shake) × 2 (hesitation: no hesitate vs. hesitate) × 2 (task performance: no mistakes vs. mistakes) between subject ANOVA was conducted on the trustworthiness ratings, and similarly for the consistency ratings. We found a main effect of task performance and motion fluency on the robot's trustworthiness. In line with our expectations, the main effect of task performance, $F(1,140) = 57.10$, $p < .001$, $\eta_p^2 = .29$ indicated that a robot which made no mistakes was rated as more trustworthy ($M = 4.59$, $SD = 1.63$) than a robot which did make mistakes ($M = 2.91$, $SD = 1.36$). The observed main effect of motion fluency revealed that the robot was trusted more when it performed its motions smoothly ($M = 4.37$, $SD = 1.71$) compared to a robot which made trembling movements ($M = 3.17$, $SD = 1.52$), $F(1,140) = 29.03$, $p < .001$, $\eta_p^2 = .17$.

The task performance × motion fluency interaction effect was also significant, $F(1,140) = 6.87$, $p = .010$, $\eta_p^2 = .05$ (see Fig. 2). Post hoc tests revealed that, when the robot made no mistakes, a trembling robot was trusted less ($M = 3.70$, $SD = 1.45$) than a smoothly moving robot ($M = 5.48$, $SD = 1.28$), $t(78) = 5.79$, $p < .001$, $r = .55$. When the robot did make mistakes, the difference between a trembling ($M = 2.61$, $SD = 1.39$) and smoothly moving robot ($M = 3.21$, $SD = 1.28$) was smaller and marginally significant, $t(74) = 2.00$, $p = .051$, $r = .22$. No other effects were significant.

5.3.3 Calibration

For the calibration of the robot's behavior in relation to its task performance, there was a significant task performance ×

[3] Autonomy ('How much did you feel the robot acted on its own') and Robot-like behavior ('How much did you feel the robot acted like you would expect from a robot') were also measured as exploratory questions. Their results can be found in the supplementary materials.

[4] A multivariate analysis of variance on all manipulation checks was also performed; apart from the expected main effects shown here, we also found some side effects. These can be found in the appendix.

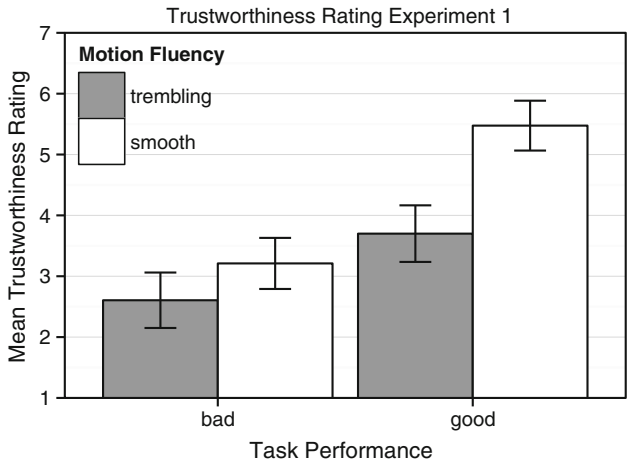

Fig. 2 Mean trustworthiness rating of the robot (*y*-axis), over performance (*x*-axis) and motion fluency (*fill*) in Experiment 1. The *error bars* represent 95% confidence intervals

motion fluency interaction effect, $F(1, 140) = 7.72$, $p < .001$, $\eta_p^2 = .05$. Post hoc t-tests revealed that when a robot made no mistakes in the task, the level of task performance was judged as less surprising when the robot was moving fluently ($M = 3.58$, $SD = 1.68$) compared to when the robot was trembling ($M = 4.55$, $SD = 1.55$), $t(78) = 2.70$, $p = .008$, $r = .29$. The difference between trembling ($M = 3.68$, $SD = 1.68$) and fluent movements ($M = 4.16$, $SD = 1.52$) was not significant when the robot made mistakes in the task, $t(74) = 1.29$, $p = .201$.

The task performance × hesitation interaction effect on calibration of performance and behavior was also significant, $F(1, 140) = 4.80$, $p = .030$, $\eta_p^2 = .03$. Post hoc tests reveal that good task performance was less surprising when the robot did not hesitate ($M = 3.58$, $SD = 1.63$) compared to when the robot did use hesitating movements ($M = 4.55$, $SD = 1.60$), $t(78) = 2.70$, $p = .009$, $r = .29$. The difference between hesitating ($M = 3.84$, $SD = 1.57$) and non-hesitating ($M = 4.00$, $SD = 1.65$) movements when the robot had bad task performance was not significant, $t(74) = .44$, $p = .663$. No other effects were observed.

5.4 Discussion

In line with the literature [14,20], the results from Experiment 1 showed that the robot's level of task performance affected judgments of its trustworthiness. We also found that motion fluency had a significant effect on the robot's judged trustworthiness in this task context. The results of this experiment show that trembling could provide an effective behavioral style to diminish trustworthiness in the Van Halen task context, both in the high and low performance conditions. Despite manipulation checks indicating that participants noticed all behavioral manipulations, gaze behavior and hesitating movements did not affect trustworthiness. Thus, we selected the motion fluency manipulation as the behavioral style manipulation in the next experiment.

For calibration, participants indicated that they found the level of performance less surprising when the robot's motion fluency (and hesitation behavior) was aligned. This indicates that participants were able to compare different sources of trustworthiness information and that behavioral manipulations, such as motion fluency and hesitation, can affect the expectations of the task performance of a robot.

The effect of motion fluency on trustworthiness was larger in the conditions in which the performance of the robot was good. This indicates a possible anchoring effect, where the baseline clip influenced the judgment of the target clip. Participants who watched a robot that made no mistakes as a target clip were subjected to a robot with the same level of task performance as in the baseline, which could have drawn more attention to the manipulations of behavioral style. Alternatively, the interaction effect might be explained as a floor effect, where a robot with bad task performance is rated as untrustworthy regardless of the level of motion fluency exhibited by the robot. This interaction effect was unexpected and we therefore have to interpret it with caution.

The results from Experiment 1 informed the design of Experiment 2, in which we used IVET to test whether the effects observed in this passive viewing task can be generalized to an interactive setting with a virtual robot.

6 Experiment 2

6.1 Introduction

In this IVE experiment, the participant and the social robot each played a Van Halen Task simultaneously. Both tasks were completely independent of one another. The manipulations that were shown to have an effect on the robot's trustworthiness in the previous experiment, task performance and motion fluency, were manipulated in a 2 (task performance: bad, good) × 2 (motion fluency: trembling, smooth) between subject design. In order to make the task more interactive, and to provide the participants with an incentive to monitor the robot, participants were tasked to correct possible mistakes the robot made by pressing a button, thereby obtaining as high a score as possible for both the robot and themselves.

6.2 Purpose

Experiment 2 served three purposes. First, it was meant to generalize our findings from Experiment 1, namely that task performance, motion fluency, and their interaction also affect a robot's trustworthiness in an interactive setting. Second, we assessed whether also participants' monitoring behavior was affected by the robot's task performance and behavioral

style. Third, we explored whether monitoring behavior is a valid measure of trust by examining its relationship with participants' trustworthiness judgments.

The robot's task performance and movement fluency were expected to have an effect on its trustworthiness, analogous to the results from Experiment 1. That is, a good task performance would make a robot more trustworthy than a bad task performance, and a smooth motion fluency would be more trustworthy than a trembling motion fluency. Also, an alignment of task performance and movement fluency was expected to be less surprising than a misalignment of these robot attributes.

By examining human robot interaction in an IVE, we are able to analyze behavioral metrics of participants' monitoring of the robot. By giving the participant an extra task to do in addition to monitoring the robot's performance, we encouraged participants to choose where to direct their attention. This can be measured by the amount of time participants spend looking at the robot. Our hypothesis was that an untrustworthy robot would recruit more attention from the participants than a trustworthy robot. This would enable the participants to potentially notice more of the robot's mistakes and correct them more quickly.

We also expected that the trust judgment as measured by the questionnaire is related to the monitoring behavior of the participants. This should be characterized as an inverse relationship, because we expect an untrustworthy robot to be monitored more than a trustworthy robot.

6.3 Method

6.3.1 Participants

87 participants (17 men, 70 women, median age: 22, age range: 18–36), recruited from the Faculty of Social Sciences of the Radboud University Nijmegen participant pool, took part in the experiment and received either course credit or a €5 gift card. Participants were randomly assigned to conditions.

6.3.2 Immersive Virtual Environment

The experiment took place in the Radboud Immersive Virtual Environment Research lab (RIVERlab). The participants wore an nVisor SX60 HMD which provided stereoscopic 3D images at a frame rate of 60 Hz in a resolution of 1280 × 1024 pixels (horizontal field of view: 44, vertical field of view: 38). On top of the HMD, and in participants' right hand, we placed sensors of the InterSense IS-900 tracking system, which were used to capture participants' head and arm movements at a sampling rate of 300 Hz. WorldViz Vizard 3.0 was used to integrate the tracker information and presentation of the virtual environment to provide an immersive experience.

6.3.3 Task

In the interactive version of the Van Halen task, participants performed their own Van Halen task concurrently with the robot while having the possibility to correct the robot when it made mistakes. Because of the layout of the virtual room (see Fig. 3), participants could not simultaneously perform their own task and monitor the robot. Participants sat on a revolving chair so their location was fixed, but they were able to look around by turning their head and/or the chair.

Scoreboards for both players were added to the task so participants could keep track of the robot's score as well as their own. Players were given a point for any non-brown ball that reaches the end of the conveyor and every brown ball picked up. For every mistake made (e.g., picking up a non-brown ball or letting a brown ball reach the end of the conveyor), the player loses one point. The scoreboards showed the amount of points gained, as well as the number of both types of mistakes made by the player.

To correct the robot, a virtual button was placed within reach of the participants. If the participants noticed the robot was making a mistake, they could correct it by pressing this button. By correcting the mistake, it would not be counted in the robot's score. However, pressing the button when the

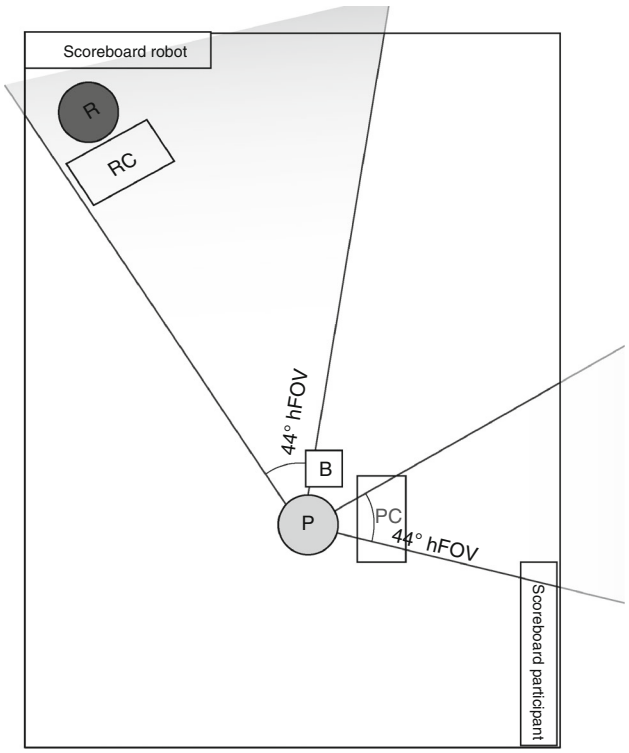

Fig. 3 Setup of the Van Halen task in Experiment 2. The robot (R) and its conveyor (RC), and the participants' conveyor (PC) are situated such that the participants (P) must rotate their head significantly to monitor the robot. The robot could be corrected by pressing the button (B)

robot made no mistake, would subtract ten points from a participant's score. This ensured that participants would not press the button randomly without monitoring the robot. The number of times participants correct the robot also measures the extent to which they direct their attention to the robot. When participants correct few mistakes, this is an indication that they either did not notice the mistakes, or felt that correcting the robot was too much effort and it made more sense to focus on their own task. The participants were instructed to gain as high as score as possible for themselves and the robot.

The task was divided in three blocks. Speed of participants' conveyor was increased in every block (Block 1: .33 m/s, Block 2: .37 m/s, Block 3: .40 m/s), as well as the number of balls generated per min (Block 1: 40 balls/min, Block 2: 46 balls/min, Block 3: 48 balls/min) to make the task incrementally more difficult. This meant the participants would have to make a deliberate choice whether to monitor the robot at the expense of their own score, or focus on their own task. The speed of the robot's conveyor was kept constant.

6.3.4 Manipulations

Task performance and Motion Fluency were manipulated in a 2 (task performance: bad, good) $\times$ 2 (motion fluency: trembling, smooth) between subject design.

In the bad task performance condition, the robot picked up correctly only half of the 16 brown balls per block. In the good task performance condition, the robot would miss two brown balls to give participants an incentive to monitor the robot in the good task performance condition as well. The robot picked up the same number of wrong balls as it missed brown balls. This approach resulted in the same number of pickup movements in both levels of task performance.

The robot's gaze behavior, examined in the previous experiment, was always turned on, since it was found to have no effect on the robot's trustworthiness rating and turning it on made the virtual environment more engaging.

6.3.5 Procedure

After obtaining informed consent form the participants, they read the Van Halen task instructions. Next, they took place on the chair and the IVE helmet and hand tracker were put on. They were then given the opportunity to look around and move their hand in the virtual environment to get accustomed to the IVE setting and notice the location of the task conveyors. After this, participants practiced their own task without the robot present to learn how to remove balls from their conveyor. Next, the robot appeared and its conveyor was started to let the participants observe the robot and practice correcting the robot's mistakes. Here, the robot would always use its fluent motions. The robot performed two mistakes, once to show a mistake in which a brown ball was not removed, and once to show a mistake where a differently colored ball was removed. The practice sessions were conducted at a low conveyor speed and generation time (.27 m/s, 30 balls/min).

After these practice sessions, there were three experimental blocks. The first 15 balls of the first block were presented at the practice speed, to accustom participants to the task. After these balls, the speed and ball generation time of both conveyors was increased. This was also the moment the robot's manipulations were activated (i.e., it started making mistakes based on its level of task performance and began moving according to its motion fluency).

Ninety balls were presented on the robot's conveyor in each block, with the exception of the first block, which was 15 balls longer because of the accustomization phase. The score of both players was reset at the beginning of each block.

After the three blocks were completed, participants filled in the questionnaire. They were then given their gift card or course credit, and dismissed.

6.3.6 Dependent Variables

Explicit Trustworthiness Rating and Calibration Participants rated the robot's trustworthiness by means of a questionnaire directly after the IVE part of the experiment was finished. Responses were recorded on a 7-point Likert scale (1= "not at all", 7 = "extremely"). Items included manipulation checks of the robot's levels of task performance and behavioral style. The explicit trustworthiness rating in this experiment is a compound measure of questions about the robot's general trustworthiness, task dependent trustworthiness, and positivity and negativity ratings. When taken together, the resulting compound variable was found to be highly reliable (Cronbach's $\alpha = .89$), which is in line with literature that suggests trustworthiness is linked with positivity judgments [44]. We decided to use this compound score in order to have a more reliable measurement than the single-item trustworthiness rating used in Experiment 1.[5] As in Experiment 1, we also included a measure about the calibration of the robot's behavior and task performance by asking how surprising the performance of the robot was based on the impression the participants got from it.[6]

[5] It is also possible to analyze Experiment 2 by means of the single item measure of trustworthiness, which yields results similar to the analysis of the compound measure reported below.

[6] As in Experiment 1, Autonomy and Robot-like behavior were also measured as exploratory questions at the end of the questionnaire. No significant effects were found for autonomy, all Fs < 1.1, all ps $> .05$. There was a significant main effect of task performance on robot-like behavior, $F(1,77) = 12.01$, $p < .001$, $\eta_p^2 = .13$, indicating that a well performing robot was judged more robot-like ($M = 4.86$, $SD = 1.41$) than a badly performing robot ($M = 3.62$, $SD = 1.79$). No other effects were significant, all Fs < 1, all ps $> .05$.

Robot Monitoring As a measure for the amount of robot monitoring, the ratio of the duration participants looked at the robot within a block over the total duration of the total effective monitoring behavior (duration of monitoring of robot plus duration of monitoring of participant task) during each block was calculated. This results in a value between 0 and 1, where a higher value indicates the robot was monitored more and a lower value indicates the robot was monitored less. For instance, when a participant has a monitoring ratio of .4 during a specific block of the experiment, he or she has looked at the robot for 40 % of the time during that block. By dividing up monitoring behavior over the three blocks of the Van Halen task, changes in monitoring behavior over time can be observed.

Number of Corrections Finally, the number of times participants correct the robot in each block was analyzed. Like the robot monitoring measure described above, the number of times participants correctly press the button to help the robot in its task can be considered an indicator of the amount of attention participants focus towards the robot.

6.4 Results

6.4.1 Exclusion Criteria

Three participants who did not complete the full three blocks were excluded (two due to time constraints and one due to nausea). Two participants who did not monitor the robot during the first block were also excluded. The behavioral data of one participant was not recorded due to a software error. This participant was also excluded from all analyses.

In total, 81 participants remained, (16 men, 65 women, median age: 22, age range: 18–36). After exclusion, each group in the 2 × 2 design consisted of 21 participants, except for the bad performance, trembling motion condition, which contained 18 participants.

6.4.2 Analyses

All explicit measures were analyzed with 2 (task performance: bad vs. good) × 2 (motion fluency: smooth vs. trembling) between subject ANOVAs. For the behavioral measures, the additional factor for ball speed (block number) was included in the analysis as a within subject factor, resulting in a 2 (task performance: bad, good) × 2 (motion fluency: smooth, trembling) as between subject factors × 3 (ball speed: slow, medium, fast) mixed design ANOVA.

6.4.3 Manipulation Checks

Manipulation checks show that participants noticed both manipulations of the robot.

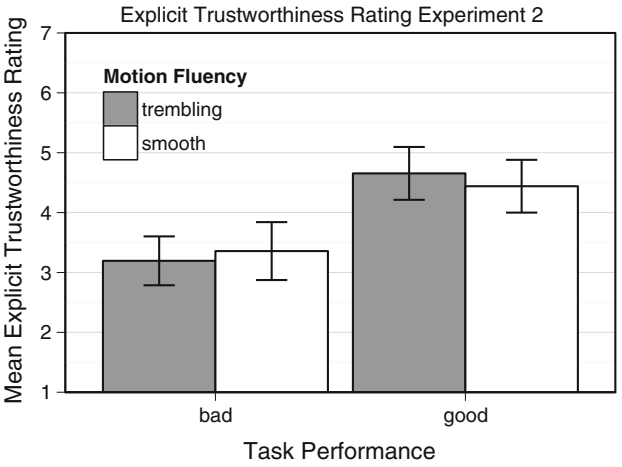

Fig. 4 Mean explicit trustworthiness ratings (*y*-axis) over task performance (*x*-axis) and motion fluency (*fill*) in Experiment 2. *Error bars* represent the 95 % confidence intervals

Task Performance Analysis of the participants' rating of robot task performance showed a significant main effect of task performance, $F(1,77) = 38.20$, $p < .001$, $\eta_p^2 = .33$. On average, participants rated the robot with good task performance higher, ($M = 4.93$, $SD = .87$) compared to a robot with bad task performance ($M = 3.49$, $SD = 1.21$). No other effects reached significance, all ps $> .05$.

Motion Fluency On average, participants rated a robot with the trembling movements as more trembling ($M = 3.97$, $SD = 1.95$) compared to a robot with fluent movements ($M = 2.05$, $SD = .96$), $F(1, 77) = 31.57$, $p < .001$, $\eta_p^2 = .29$. No other effects reached significance, all ps $> .05$.

6.4.4 Explicit Trustworthiness Rating and Calibration

The analysis of the explicit trustworthiness rating revealed a main effect of task performance, $F(1,77) = 35.09$, $p < .001$, $\eta_p^2 = .31$. On average, a robot with bad performance is trusted less ($M = 3.28$, $SD = .95$) than a robot with good performance ($M = 4.55$, $SD = .96$). No effects were found for motion fluency and task performance × motion fluency interaction. See also Fig. 4.

No effects were found on the measure for calibration of the robot's behavior and performance, including the expected task performance × motion fluency interaction effect, $F < 1$.

6.4.5 Monitoring

Analysis of the robot monitoring behavior by participants showed that the main effect of task performance was significant, $F(1, 77) = 7.78$, $p = .007$, $\eta^2 = .07$, such that a robot with bad performance was monitored more ($M = .36$,

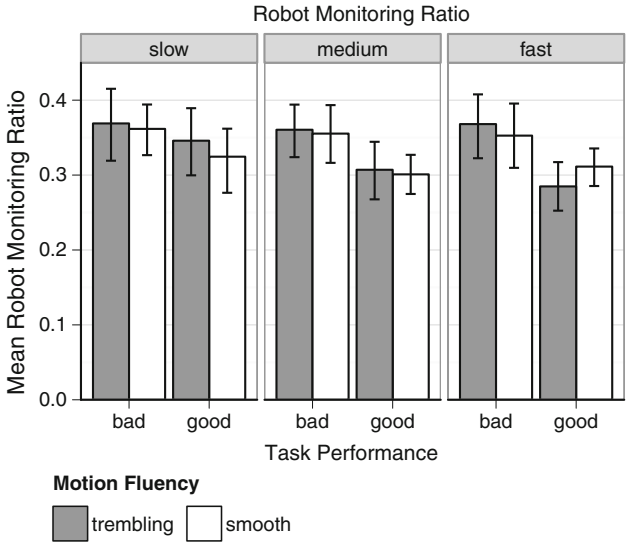

Fig. 5 Mean ratio of robot monitoring over different participant conveyor speeds (*panels*), robot task performance (*y*-axis), and motion fluency (*fill*). The *error bars* represent 95 % confidence intervals

$SD = .09$) than a robot with good performance ($M = .31$, $SD = .08$).

There also was a significant main effect of ball speed, $F(2,154) = 4.65$, $p = .011$, $\eta^2 = .01$. Planned contrasts revealed that during the slow ball speed, the robot was monitored more ($M = .35$, $SD = .10$) than during the medium ($M = .33$, $SD = .09$) and fast ($M = .33$, $SD = .09$) speeds, $t(154) = 3.04$, $p < .003$. The effects of medium and fast speeds did not differ significantly from each other, $t(154) = .23$. No other effects were significant (all ps $> .05$) (see Fig. 5).

6.4.6 Relationship Between Explicit Trust Rating and Monitoring

From the convergent patterns observed for the explicit trust ratings and monitoring behavior, it seems likely that a relationship in the form of correlations or mediation effects between the two measures exist. However, no significant relationship was observed. Thus, we did not find evidence that participants' trustworthiness judgments are related to their monitoring behavior.

6.4.7 Number of Corrections

Finally, an exploratory analysis on the number of times the robot was corrected was carried out. The analysis of the number of corrections metric showed the main effect of task performance to be significant, $F(1, 77) = 248.95$, $p < .001$, $\eta^2 = .73$. A robot in the good performance condition was corrected less ($M = 2.04$, $SD = .97$) than a robot in the bad task performance condition ($M = 9.54$, $SD = 3.30$). Interestingly, the task performance × motion fluency × ball speed 3-way interaction effect was marginally significant, $F(2, 154) = 2.55$, $p = .081$, $\eta^2 = .01$, which upon further examination was driven by a marginally significant 2-way task performance × motion fluency interaction during the medium ball speed, $F(1, 77) = 6.49$, $p = .070$, $\eta^2 = .08$. Post hoc t tests on the medium speed block reveal, aside from the previously described main effect of task performance, that there is a significant difference between trembling motions ($M = 10.94$, $SD = 3.17$) and fluent motions ($M = 8.81$, $SD = 2.98$) when the robot is performing badly, $t(37) = 2.17$, $p = .037$, while the difference between the motion fluency conditions in the good task performance condition is not significant, $t(40) = 1.42$, $p = .162$.

6.5 Discussion

In this Experiment, a robot which has good task performance was shown to be trusted more than a robot with bad task performance on the explicit trustworthiness rating. Moreover, a robot with good task performance was monitored less than a robot with bad task performance. Although manipulation checks show that participants noticed the motion fluency manipulation, no evidence was found for the expected effects of motion fluency on the explicit trustworthiness rating and monitoring. Moreover, the expected task performance × motion fluency interaction effect on the calibration of the robot's motion fluency and task performance was not found. An analysis of the number of corrections did seem to reveal an effect of motion fluency when the robot had bad task performance, during the medium speed block of the experiment. Although we must be careful interpreting results from such an exploratory analysis, it does give an indication we are on the right track here.

Contrary to our expectations, the explicit trust rating was not related to the monitoring behavior of the participants. This indicates that although the performance manipulation did have an effect on both the explicit trust rating and the monitoring behavior, participants did not seem to base their explicit trust on their monitoring of the robot. Conversely, it can be said that a low propensity to trust in the Van Halen task does not seem to directly influence participants' monitoring behavior during the task.

A possible explanation for the absence of this effect is that the setup of the collaborative task in this experiment may have been too constraining on the monitoring behavior of the participants. There may have been optimal monitoring strategies for either levels of the robot's task performance, and little room for spontaneous, trust related monitoring. For instance, the most effective strategy could have been to look at the robot when participants' own conveyor did not have brown balls on it. Consequently, too little spontaneous behav-

ior related to trust might have been present in the monitoring behavior of the participants.

These results from Experiment 2 are not completely in line with the findings of Experiment 1, in which motion fluency did have an effect on the robot's trustworthiness and a significant task performance × motion fluency interaction was found. It is possible that one or more differences between the video study conducted in Experiment 1 and the IVE study in Experiment 2 caused the absence of a motion fluency effect on the robot's trustworthiness.

One explanation might be that participants were too focused on their own conveyor to connect the motion fluency of the robot to its trustworthiness. Alternatively, the length of the experiment might have influenced the outcome. Participants may have grown accustomed to the motion fluency of the robot. While motion fluency may have had an effect on the trustworthiness of the robot early in the experiment, participants were asked to rate the robot much later, at the end of the experiment. Within this time, a reasonable estimate of the robot's task performance can been created and participants may have discarded the trustworthiness information derived from the robot's motion fluency. Given that previous findings in social psychology indicate that the effect of appearance is still present after multiple observations of task performance [20,45], it future research may investigate whether these findings do or do not apply to prolonged observations of behavioral style. Another difference was the presence of a baseline video in Experiment 1, which was not present in Experiment 2. Although manipulation checks show that the participants noticed the motion fluency manipulation in both experiments, the baseline video in Experiment 1 may have drawn more attention to this manipulation. Finally, individual expectations about robots in general, and the propensity to trust a robot will likely influence a participant's perception of a robot and its trustworthiness. In future research, it would be valuable to measure these variables as pre-measures in order to take these idiosyncratic differences into account.

It is also worth pointing out that a single behavioral feature, such as motion fluency, may in itself contain insufficient information to influence an explicit trustworthiness judgment in an interactive setting. Indeed, recent work on economic exchange after a short conversation between strangers has shown that trustworthiness is derived from a multitude of small nonverbal gestures [15]. It is possible that a similar, more encompassing strategy to use such behavioral features for trustworthiness judgments might work here as well.

7 Summary and Conclusion

As far as we know, the current research is the first to combine both performance and behavioral manipulations of trustworthiness of a humanoid robot in an HRI scenario. In the experiments presented here, we have reaffirmed that a social robot's trustworthiness is chiefly influenced by its performance on a task [14]. Features of a robot's behavioral style (such as motion fluency, hesitations and gaze behavior in Experiment 1), and motion fluency (in Experiment 2) are noticed by participants. There is also preliminary evidence these manipulations may lead to changes in the participants' behavior, although we should not jump to conclusions, considering the exploratory nature of this analysis. Motion fluency was identified to affect a robot's trustworthiness in a non-collaborative VHRI experiment. In the follow up experiment in an interactive setting using IVET, this effect was not present. In the interactive setting, a robot's task performance did seem to have a marginal significant influence on participants' monitoring behavior during the experiment but it was not possible to connect monitoring behavior with the explicit trustworthiness judgment about the robot.

The current research also demonstrates that it is possible to experimentally investigate HRI with video and IVET. As we have noted, all methods used in social HRI research have their merits and downsides. Therefore, researchers might first want to try out the effects of their designs in a relatively inexpensive video experiment to see whether robot behavior designed in simulation generates the expected explicit response in participants. If these results are satisfactory, interactive experiments in IVE can be conducted with a virtual robot, and as a final step the results from the IVE experiments should be validated in a dynamic live robot experiment. Studies using VHRI and IVE paradigms do permit a relatively efficient way of manipulating robot behavioral styles, and sampling human responses to them. For instance, if Experiment 2 had been carried out with a real robot instead, the cost would have been much higher than with the IVE that we used. From this perspective, video and IVET studies should not be seen as the end goal in a research line, but can best be treated as valuable stepping stones to explore the feasibility of the many potential avenues in Human Robot Interaction research.

This research also reveals the difficulties involved in replicating the effects found in a non-interactive video experiment to more interactive settings using IVET, and it is to be expected that results from IVET studies in turn do not directly translate to HRI in the real world. It is possible that effects of behavioral style on trust might be found more easily when participants can direct their complete attention to the robot, while in more realistic settings participants can get too distracted to make these higher level inferences about a robot's behavior. The fact that the effects from a subtle behavioral manipulation did not carry over from a relatively passive experimental paradigm (VHRI) to another more interactive one (IVE) is an important consideration when designing future social HRI experiments.

References

1. Young JE, Hawkins R, Sharlin E, Igarashi T (2009) Toward acceptable domestic obots: applying insights from social psychology. Int J Soc Robot 1:95–108
2. Sztompka P (1999) Trust: a social theory. Cambridge University Press, Cambridge
3. Sanfey A (2007) Social decision-making: insights from game theory and neuroscience. Science 318:598–602
4. Mayer R, Davis J, Schoorman F (1995) An integrative model of organizational trust. Acad Manag Rev 20(3):709–734
5. Schoorman F, Mayer R, Davis J (2007) An integrative model of organizational trust: past, present, and future. Acad Manag Rev 32(2):344–354
6. Simpson JA (2007) Psychological foundations of trust. Curr Dir Psychol Sci 16(5):264–268
7. Lee J, See K (2004) Trust in automation: designing for appropriate reliance. Hum Factors 46(1):50–80
8. Axelrod R, Hamilton WD (1981) The evolution of cooperation. Science 211(4491):1390–1396
9. Willis J, Todorov A (2006) First impressions: making up your mind after a 100-Ms exposure to a face. Psychol Sci 17(7):592–598
10. Oosterhof NN, Todorov A (2008) The functional basis of face evaluation. Proc Natl Acad Sci 105(32):11087–11092
11. Dotsch R, Todorov A (2012) Reverse correlating social face perception. Soc Psychol Pers Sci 3(5):562–571. doi:10.1177/1948550611430272 NULL
12. Kaul TJ, Schmidt LD (1971) Dimensions of interviewer trustworthiness. J Couns Psychol 18(6):542–548. doi:10.1037/h0031748
13. Roll WV, D SL, Kaul TJ (1972) Perceived interviewer trustworthiness among black and white convicts. J Couns Psychol 19(6):537–541
14. Hancock PA, Billings DR, Schaefer KE, Chen JYC, de Visser EJ, Parasuraman R (2011) A meta-analysis of factors affecting trust in human–robot interaction. Hum Factors 53(5):517–527
15. DeSteno D, Breazeal C, Frank RH, Pizarro D, Baumann J, Dickens L, Lee JJ (2012) Detecting the trustworthiness of novel partners in economic exchange. Psychol Sci 20(10):1–8. doi:10.1177/0956797612448793
16. Petty RE, Cacioppo JT (1986) Communication and persuasion: central and peripheral routes to attitude change. Springer, New York
17. Brewer MB (1988) A dual process model of impression formation. Erlbaum Associates, Hillsdale
18. Berg J, Dickhaut J, McCabe K (1995) Trust, reciprocity, and social history. Games Econ Behav 10(1):122–142
19. van 't Wout M (2008) Friend or foe: the effect of implicit trustworthiness judgments in social decision-making. Cognition 108:796–803
20. Chang LJ, Doll BB, van 't Wout M, Frank MJ (2010) Seeing is believing: trustworthiness as a dynamic belief. Cogn Psycho 61:87–105
21. Mumm J, Mutlu B (2009) Human–robot proxemics: physical and psychological distancing in human–robot interaction. In: Proceedings of artificial intelligence and simulation of behavior convention (AISB 09), Lausanne, Switzerland
22. Takayama L, Pantofaru C (2009) Influences on proxemic behaviors in human–robot interaction. In: Proceedings of the IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, St Louis, Missouri, pp 5495–5502
23. Heerink M, Kröse B, Evers V, Wielinga B (2010) Relating conversational expressiveness to social presence and acceptance of an assistive social robot. Virtual Real 14:77–84
24. Walters ML, Lohse M, Hanheide M, Wrede B, Syrdal DS, Koay KL, Green A, Hüttenrauch H, Dautenhahn K, Sagerer G (2011) Evaluating the robot personality and verbal behavior of domestic robots using video-based studies. Adv Robot 25:2233–2254
25. Walters ML (2008) The design space for robot appearance and behaviour for social robot companions. PhD Thesis, University of Hertfordshire
26. Syrdal DS, Koay KL, Gácsi M, Walters ML, Dautenhahn K (2010) Video prototyping of dog-inspired non-verbal affective communication for an appearance constrained robot. In: Proceedings of the 19th IEEE international symposium on robot and human interative communication. Principe de Piemonte, Italy, pp 632–637
27. Syrdal DS, Otero N, Dautenhahn K (2008) Video prototyping in human–robot interaction: results from a qualitative study. In: Abascal J, Fajardo I, Oakley I (eds) Proceedings of the 15th European congerence on cognitive ergonomics: the ergonomics of cool interaction. ACM New York, NY, Madeira, Portugal, pp 1–8
28. Takayama L, Dooley D, Ju W (2011) Expressing thought: improving robot readability with animation principles. In: Proceedings of human–robot interaction conference: HRI 2011. Lausanne, Switzerland, pp 69–76
29. Dautenhahn K (2007) Methodology & themes of human–robot interaction: a growing research field. Int J Adv Robot Syst 4(1):103–108
30. Blascovich J, Loomis J, Beall A, Swinth K, Hoyt C (2002) Immersive virtual environment technology as a research tool for social psychology. Psychol Inq 13(2):103–124
31. Groom CJ, Sherman JW, Conrey FR (2002) What immersive virtual environments can offer to social cognition. Psychol Inq 13(2):125–128
32. Dotsch R, Wigboldus DHJ (2008) Virtual prejudice. J Exp Soc Psychol 44:1194–1198
33. Rinck M, Rörtgen T, Lange WG, Dotsch R, Wigboldus DHJ, Becker ES (2010) Social anxiety predicts avoidance behaviour in virtual encounters. Cogn & Emot 24(7):1269–1276. doi:10.1080/02699930903309268
34. Rinck M, Kwakkenbos L, Dotsch R, Wigboldus DHJ, Becker ES (2010) Attentional and behavioural responses of spider fearfuls to virtual spiders. Cogn & Emot 24(7):1199–1206. doi:10.1080/02699930903135945
35. Tikhanoff V, Cangelosi A, Metta G (2011) Integration of speech and action in humanoid robots: iCub simulation experiments. IEEE Trans Auton Ment Dev 3(1):17–29
36. Woods S, Walters M, Koay KL, Dautenhahn K (2006) Comparing human robot interaction scenarios using live and video based methods: towards a novel methodological approach. In: Proceedings of the 9th IEEE international workshop on advanced motion control (AMC'06), New York. IEEE Press, Istanbul, Turkey, NY, pp 750–755
37. 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
38. Bagheri N, Jamieson GA (2004) Considering subjective trust and montioring behavior in assessing automation-induced "Complacency". In: Proceedings of the human performance, situation awareness and automation conference, SA Technologies, Marietta, GA, pp 1–6
39. Walters ML, Dautenhahn K, Te Boekhorst R, Koay KL, Syrdal DS, Nehaniv CL (2009) An empirical framework for human–robot proxemics. In: New frontiers in human–robot interaction, Edinburgh, Scotland
40. Iwata H, Sugano S (2009) Design of human symbiotic robot TWENDY-ONE. In: IEEE international conference on robotics and automation, pp 580–586
41. Ambady N, Weisbuch M (2010) Nonverbal behavior. Handbook of social psychology. Harvard University Press, New York, pp 464–497

42. Emery NJ (2000) The eyes have it: the neuroethology, function and evolution of social gaze. Neurosci Biobehav Rev 24(6):581–604
43. Srinivasan V, Murphy R (2011) A survey of social gaze. Human–robot interaction (HRI). Lausanne, Switzerland, pp 253–254
44. Todorov A (2008) Evaluating faces on trustworthiness: an extension of systems for recognition of emotions signaling approach/avoidance behaviors. Ann N Y Acad Sci 1124(1):208–224
45. Delgado MR, Frank RH, Phelps EA (2005) Perceptions of moral character modulate the neural systems of reward during the trust game. Nat Neurosci 8(11):1611–1618

Rik van den Brule obtained his M.Sc degree in Artificial Intelligence at Radboud University Nijmegen, and is currently a junior researcher at the Donders Institute for Brain, Cognition and Behaviour and the Behavioural Science Institute. His research focuses on trust in Human-Robot Interaction.

Ron Dotsch is assistant professor at the Behavioural Science Institute, Radboud University Nijmegen. His research interests include perception of social signals, cooperation, person perception and social cognition using data-driven research methods.

Gijsbert Bijlstra is an assistant professor in Work and Organizational Psychology at the Behavioural Science Institute of the Radboud University Nijmegen. His research is focused on the perception of nonverbal behavior in social interactions.

Daniel H. J. Wigboldus is professor of social psychology at the Behavioural Science Institute of the Radboud University Nijmegen. His main research interest is person perception, with a focus on stereotyping, prejudice and face perception.

Pim Haselager is a principal investigator of the Donders Institute for Brain, Cognition, and Behaviour and a staff member of the Department of Artificial Intelligence at the Radboud University Nijmegen, The Netherlands. His research interests include human–robot interaction, and the ethical, legal and societal implications of robotics.