

A SHAKY FOUNDATION FOR TRUST
EFFECTS OF TASK PERFORMANCE AND MOVEMENT
STYLE ON TRUST AND BEHAVIOR IN SOCIAL
HUMAN-ROBOT INTERACTION

Author
R. van den Brule (0425842)

Supervisors
Pim Haselager Donders Center for Cognition
Ron Dotsch Behavioral Science Institute

July 14, 2010

Radboud University Nijmegen



DEPARTMENT OF ARTIFICIAL INTELLIGENCE

MASTER THESIS IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF SCIENCE IN ARTIFICIAL INTELLIGENCE

Abstract

Consumer robots are slowly beginning to emerge as household appliances. As these robots become more sophisticated and are treated more as an addition to the family, it is important to equip them with the social skills needed to be accepted and trusted in a household environment. The Trust Me project aims to develop a social robot which can calibrate its trustworthiness based on the behavior of the person it is interacting with.

Trust is an important factor in social interaction, as it is the attitude towards an agent in which it will help achieve an individual's goal. Task performance of the agent is seen as the most objective way to estimate its trustworthiness, but is difficult to observe before interaction takes place. Therefore, the agent's appearance and behavioral style (e.g. movements, display envelopes) is commonly used as a way to assess its trustworthiness. However, the agent's task performance and appearance do not have to correlate.

The relationships between trust, behavior, task performance and appearance and behavioral style are not well understood. An experiment was performed in an Immersive Virtual Environment (IVE) in which participants had to perform a social decision task with a robot avatar. The task performance and movement style of the robot avatar were manipulated. The robot could have either a bad or good task performance and a shaky or smooth movement style.

Results show that robots with a better task performance are generally trusted more than robots with bad task performance. At the same time, robots which have their level of movement style aligned with their level of task performance are trusted more than robots which have inconsistent levels of movement style and task performance. These results suggest that while it is important for a social robot to perform well on a task in order to be trusted, it is also important to show uncertainty by altering its behavioral style when the robot cannot perform a task satisfactorily.

The fit effect of trust is also found in participant reaction times, distances kept from the robot, and movement speeds. The trust variables mediate these behavioral metrics. In future research on-line measurement of these metrics can be used by the robot to estimate its own trustworthiness. Potentially, the robot can utilize this information to alter its behavioral style to evoke the right amount of trust from its user.

Acknowledgements

This thesis is the result of almost a year of literature studies, experiment development, data analysis and brainstorm sessions. As a result, many people have influenced me or had a part in the result. At this point I would like to thank all of them for their contributions. Pim Haselager for his supervision and creative input. Jop van Heesch for coming up with the experimental task. Marlon Baeten, with whom Jop and I wrote all the software for the Experiment in the Interactive Virtual Environment. Jeroen Derks for creating the beautiful robot avatar. Gijs Bijlstra for creating the questionnaire and his help with statistics. Ron Dotsch, also for his help with the statistical analyses, his suggestions for behavioral metrics, general tips about the Virtual Environment, and his explanation of mediation. Eva Cillessen for modeling as a participant in the Virtual Reality lab. Gea Dreschler of the Academic Writing Center who helped me edit this thesis. Bastiaan du Pau for our common clashes with and eventual triumph over L^AT_EX. All the other people with whom I shared the “stagekamer” and who had to put up with my clutter.

Contents

Abstract	i
Acknowledgements	iii
1 Introduction	1
1.1 Robots in Real-world Environments	1
1.2 Social Robotics	2
1.3 Trust	2
1.4 The Trust Me Project	5
1.5 Research question and hypotheses	5
1.6 Interactive Virtual Environment	6
1.7 Robot	6
1.8 Experiment	7
2 Method	9
2.1 Participants	9
2.2 Task	9
2.3 Conditions	10
2.4 Setup of the Immersive Virtual Environment	10
2.5 Metrics	11
2.6 Experiment Procedure	13
3 Results	15
3.1 Participants	15
3.2 Test Information	15
3.3 Explicit Trust	15
3.4 Behavioral metrics	18
3.5 Correlation and Mediation of Trust on behavior	21
4 Discussion	25
4.1 Summary of results	25
4.2 Effects on Trust	25
4.3 Role of Trust on Behavior	26
4.4 Conclusions	27
5 Further Research	29
5.1 Aspects of Appearance and Behavioral Style	29
5.2 Change of trust over time	29
5.3 Looking behavior and neglect time	30
5.4 Realism	30

References	33
Appendices	35
A Questionnaire	37
B Mediator Regression Models	39
B.1 Mediation of Affect Based Trust on Behavior	39
B.2 Mediation of Cognition Based Trust on Behavior	39
B.3 Mediation of total trust	39
C Results for all Participants	43
C.1 Differences between Dutch and German participants	43
C.2 Analysis of Difference	43

Chapter 1

Introduction

The Trust Me research project is aimed at studying perceived trust in social human-robot interaction. The project uses immersive virtual environment (IVE) technology to study human-robot interaction.

In this Chapter, overviews of robots in real world environments (Section 1.1) and Social Robotics (Section 1.2) are presented. Next, an overview of Trust is given in Section 1.3 and the purpose of the Trust Me project is further explained in Section 1.4. The research questions of this thesis are presented in Section 1.5. Then, the Immersive Virtual Environment is explained in Section 1.6. The design of our own virtual robot is presented in Section 1.7. The social task used in the experiment is explained in Section 1.8.

1.1 Robots in Real-world Environments

For most part of the second half of the 20th century, robots have primarily existed in factories where production has become more and more automated. In the past decade, a new brand of robots has begun to emerge – robots for ordinary people, to be used in their households. These so-called ‘consumer robots’ are used purely as a source of entertainment, such as Sony’s Aibo, but functional robots such as robotic vacuum cleaners (iRobot’s Roomba) and lawn mowers have become available as well. There is ongoing research into humanoid robot assistants for elderly people (e.g. TWENDY-one (Iwata & Sugano, 2009)). This trend is likely to continue, and this class of robots is expected to become more common in our society in the near future as robots for the home environment become more available.

In a household setting much more interaction between robots and humans takes place. The main difference between a factory and a household is the level of control – where factories usually are highly controlled environments, homes are usually not set up like an assembly line. Robots that operate in a household setting not only need to be able to interact with the world around them, but also with the people in it. This requires a higher level of autonomy. Apart from that, robots in a household environment are co-located with humans. Whereas a construction robot in a factory may be physically separated from humans, household robots occupy the same rooms as their owners. This is a distinction between remote and proximate interaction (Goodrich & Schultz, 2007). Proximate interaction will not only result in more frequent interactions but also in interactions that are more social in nature.

Furthermore, people tend to attribute more human qualities to robots than they actually have. Even a simple autonomous vacuum cleaner such as the Roomba can

have a profound effect on family life and is sometimes even given a name (Forlizzi, 2007). It would seem that an autonomous vacuum cleaner is treated more like a household pet instead of an appliance while its functionality and appearance clearly were not designed with that in mind. If such a robot, which has no physical resemblance of a human or pet at all, is already treated as a social agent, the question arises what would happen when humanoid robots are introduced. The automatic desire to anthropomorphize with such robots will most likely only be larger.

If such robots were equipped with some form of social skills, it could make it easier for people to interact with them and make predictions about their behavior. A socially interactive robot should utilize the people's need to anthropomorphize with it to facilitate better (social) interaction. Research into this field is already under way, and the state of this field of Social Robotics is summarized in Section 1.2.

1.2 Social Robotics

Socially interactive robots can be broadly defined as “robots for which social human-robot interaction is important” (Fong, Nourbakhsh, & Dautenhahn, 2003). Consumer robotics fall under this category because of their co-locatedness with humans and the social nature of their environment. If consumer robots should be given the social skills it needs to function properly in our social world, it is worth while to take a look at this research field.

In recent years, research has increased in this field. Results of such research projects include Kismet (Breazeal, 2002, 2003) and Sparky (Scheef, Pinto, Rahardja, Snibbe, & Tow, 2002). Interacting with people and objects in a home setting is a quite complex process. Because of this, the robots that result from these projects are either very sophisticated (Kismet uses a network of 10 computers for its visual system alone), or there is a ‘human in the loop’ who acts as the decision maker, as is the case with Sparky. Both approaches have their downsides. In the case of Kismet, much time and effort is spent on the development of control systems before interaction can take place, while using teleoperation decreases autonomy, as is the case with Sparky. To experiment with social interaction on a large scale, a system simpler than Kismet yet more autonomous than Sparky is required.

1.3 Trust

As it is for all social interactions, trust is an important factor in social human-robot relations. This is especially the case when a robot is expected to perform certain tasks for its owner. Lee and See (2004, pp. 55) define trust as “the attitude that an agent [automation, robot, or human] will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability”.

1.3.1 The Trust Relationship

In a trust relationship, the ‘trustee’ is the party which has to be trusted, and a ‘trustor’ is the party which has to trust the trustee. The trustor needs to interact with a trustee to achieve a particular goal. The question is how the trustor can find out how well the trustee can be trusted in helping the trustor achieve his goal. In other words, the trustor has to estimate the trustworthiness of the trustee.

There is an important distinction between the *perceived* trustworthiness of the trustee by the trustor and the *actual* trustworthiness of the trustee. Objectively, the

trustee should only be trusted as much as his or her capability of performing a certain task effectively and reliably. However, the amount of trust a trustor genuinely places in the trustee is, apart from the trustee’s task performance, also based on the trustee’s appearance, behavioral and cognitive capacities displayed to the trustor.¹ These aspects of the trustee are more easily observable than its task related performance behavior because they can be discovered by simply looking at the trustee – whereas to observe the trustee’s task performance, one needs to have completed a trust interaction at least once with the trustee.

The use of appearance and presentation in the assessment of trustworthiness is useful when the trustee and trustor have not had a previous trust relationship and an objective trustworthiness assessment cannot be made on the basis of the trustee’s task performance. In that case, the presentation of the trustee is everything, and hopefully the way the trustee presents himself is calibrated with his task performance capabilities. Assuming that the easily observable appearance aspects of a trustee corresponds with its as of yet unobserved task performance, appearance and behavior give a trustor the means to estimate the trustworthiness of the trustee.

Unfortunately, it is also possible that the trustee’s presentation of itself is not aligned with its task performance. For instance, a con man tries to win over people to hand over their money to him by presenting himself as a respectable business man. Sadly for his victims, the con man runs away with their ‘investments’ at their expense.

This discrepancy of low task performance and high appearance and presentation leads to overtrust of the trustee. Similarly, high task performance coupled with a bad quality of presentation and appearance leads to undertrust of the trustee (see also Table 1.1).

		Actual Trustworthiness	
		Low	High
Perceived Trustworthiness	High	Overtrust	Calibrated (high)
	Low	Calibrated (low)	Undertrust

Table 1.1: Possible outcomes in calibrating actual and perceived trustworthiness of the trustee for the trustor, when its task performance has not yet been observed by the trustor.

When the trustor has had the opportunity to observe the task related performance of the trustee by interacting with him, task related performance of the trustee will start to affect the trustor’s perceived trustworthiness of the trustee. The question is what the influence of this performance related effect is on the perceived trustworthiness. In a purely additive model, trustworthiness from the perceived task performance will add to the trustworthiness generated by its perceived appearance of the trustee. In that case, the trustee with good task performance and a good presentation and appearance would be trusted the most by the trustor, while a trustee which has both bad task performance and a bad appearance and presentation would be trusted the least.

However, it is also possible that trustees which present and act *consistently* are better trusted than trustees who act *inconsistently* of the way they present themselves,

¹That appearance influences attitude towards people has been established by Clifford and Walster (1973). Teachers were asked to estimate a child’s intelligence and popularity based on a report card, accompanied with a picture of an attractive or unattractive child. Teachers gave higher ratings when a picture of an attractive child was attached to the report card than when a picture of an unattractive child was shown with the report card.

as in the case with the con man example. This is also called *goodness of fit*. Both models are illustrated in Figure 1.1.

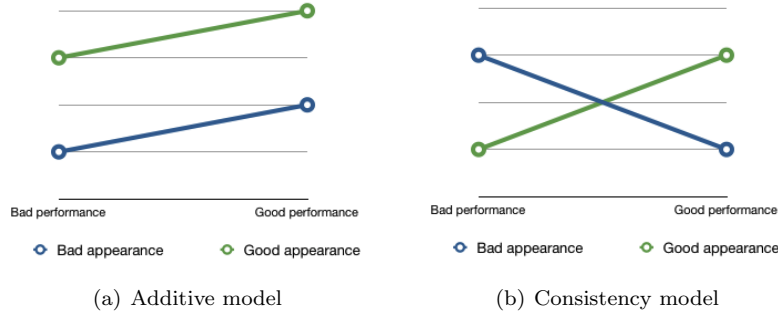


Figure 1.1: Two models of perceived trustworthiness of the trustee by a trustor (y -axis), in relation to its task related performance behavior (x -axis), and its appearance and presentation (separate lines). The additive model (a) assumes that the relationship between task performance and appearance is additive, while the consistency model (b) assumes trust is highest when task performance and appearance are consistent with each other.

Assuming the appearance and presentation of the trustee remains constant, the trustworthiness of the trustee will change over time as the trustor observes more and more task related performance behavior of the trustee.

In short, perceived trustworthiness is influenced by two observable aspects of the trustee’s behavior. First, the trustee has task related performance behavior. Second, the trustee has a specific appearance and behavioral style (ABS), which consists of all its non-performance related aspects. For a social robot, examples of ABS are properties such as the size, body shape, and movement style of the robot. The robot’s actual trustworthiness is solely influenced by its task related performance behavior. A robot’s perceived trustworthiness is mostly determined by its task related performance behavior, but its ABS is likely to be an important contributing factor as well.

1.3.2 Measuring Trust

Trust can be measured in two ways. The first way to measure trust is to ask the trustor about his or her attitude towards the trustee. This is an explicit metric, because the trustor needs to put his trust attitude into words. In an experimental setting, a questionnaire is usually used to derive an explicit trust score. Madsen and Gregor (2000) developed such a questionnaire for human-computer trust. They divide trust into different constructs which are either cognition based or affect based. Cognition based trust constructs are questions like “My partner has the same goals I do”, while affect based constructs are questions such as “I like my partner”.

The second way to measure trust is to look at aspects of the trustor’s behavior which correlate with his or her trust attitude. Behavior can be observed implicitly. However, very little is known about which behaviors correlate with trust attitude. In human-robot trust, the metric most often used is the robot’s Neglect Time or Neglect Tolerance. This metric measures the time a robot can be left unattended by his supervisor or operator (Steinfeld et al., 2006).

1.4 The Trust Me Project

The aim of the Trust Me Project is to study social human-robot interaction. Ultimately, the goal is to develop a social robot which can assess his own trustworthiness on the basis of the behavior of the person it is interacting with (Figure 1.2). Such a robot could then use this information to alter its ABS to re-calibrate his trustworthiness so that no undertrust or overtrust occurs (See Table 1.1).

The Trust Me project uses a model of trust which divides the robot's perceived trustworthiness into two different categories: explicit and implicit trust. Explicit trust refers to the conscious experience of trust by the user. It is open for introspection and the participant can be asked about his or her explicit trust in the robot (e.g. by filling out a questionnaire). Implicit trust refers to the automatic, involuntary processes which cause people to attribute a level of trust quickly, as opposed to the slower and deliberate explicit trust system. This process is not conscious, so in order to measure its effect, other metrics need to be developed related to the behavior of the participant.

Unfortunately the behavioral correlates of trust are not well known. Before an attempt of constructing a robot which can calibrate the trust of its human companion can be undertaken, these relationships have to be discovered. Therefore, we performed an exploratory study in which a participant performed a social decision task with a social robot in a virtual environment. The robot's task related performance behavior and its behavioral style were manipulated to measure the effects on its perceived trustworthiness.

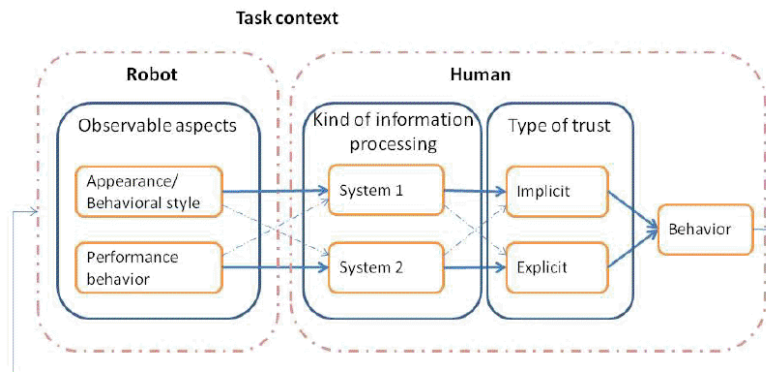


Figure 1.2: The model of trust processing as proposed by Haselager et al. (2009). The Appearance and Behavioral Style (ABS) is mostly processed by a low-level cognitive system (System 1) and leads mainly to implicit trust. Task related Performance Behavior is mostly processed by a high-level cognitive system (System 2) and leads mainly to explicit trust. Trust leads to certain behavior in the trustor, which the robot (the trustee) can use to measure its trustworthiness.

1.5 Research question and hypotheses

The main research questions of this thesis are:

1. What is the effect of the task related performance behavior of an interactive social robot

- on its perceived trustworthiness by the participant?
 - on the behavior of the participant?
2. What is the effect of the behavioral style of an interactive social robot
 - on its perceived trustworthiness by the participant?
 - on the behavior of the participant?
 3. Can the changes in the behavior of the participant be explained by changes in their perceived trustworthiness?

1.6 Interactive Virtual Environment

Immersive Virtual Environments (IVEs) are becoming widely used in the scientific community. In the field of social psychology, IVEs are used as methodological tools for experiments because they make it possible to create settings which are both controlled *and* realistic, enabling easier replication of experiments (Blascovich et al., 2002). Also, using an IVE it is easier to measure human behavior continuously and unobtrusively, because IVE keeps track of the user's movements. A good example of the use of IVE is an experiment by Dotsch and Wigboldus (2008), where the distance the participant keeps between himself and avatars of different ethnic appearances is used as a measurement of implicit associations with ethnicity.

IVEs can be used as a research platform for Social Robotics as well. Real Social Robots are still very expensive to create and maintain. However, by creating them in a virtual environment, many costs can be reduced. A virtual representation or avatar of a social robot is much easier to create than a real social robot. Also, many of the problems in (social) robotics described in Section 1.2 are easier to solve with the use of an IVE. Everything in a virtual world is formally stored in computer memory. This information can be accessed by a virtual robot, which makes the implementation of the robot's positioning, object location, eye contact, and visual attention systems much simpler.

1.7 Robot

A virtual robot was designed for Human-Robot interaction at Radboud University Nijmegen. Both the appearance of the robot and the behavioral and movement system were developed in-house.

An important aspect of the robot appearance was that it needed to look like a robot, but also enough like a human to induce anthropomorphism. When a robot is made more humanlike in appearance and motion, the response of a human to the robot becomes more positive, until a point is reached when the positive response quickly changes to discomfort. When the appearance and movements of robots become indistinguishable from a human, the empathic response towards it becomes positive again (Mori, 1970). This effect is called the *uncanny valley*. Because of this effect, robots are usually built more to look like a machine than is possible with current technology, thus placing them on the first peak before the valley.

We decided to base the appearance of our robot on TWENDY-one, a recent prototype of a robotic assistant for elderly people (Iwata & Sugano, 2009). This robot looks robot-like but also resembles a human enough to evoke an empathic response (see Figure 1.3).

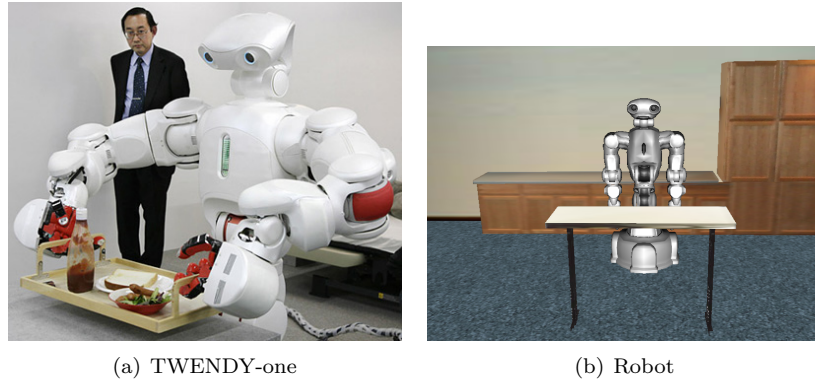


Figure 1.3: Our inspiration for the appearance of our social robot, TWENDY-one (Iwata & Sugano, 2009) (a), and the virtual robot used in the experiment (b).

The virtual robot should be able to initiate and respond to social interactions. Humans use paralinguistic social cues, also called *display envelopes* (Breazeal, 2003), to regulate speaking turns. The robot also needed to respond to cues given by the user such as gestures and should also be able to attract the user’s attention by making gestures or some vocalizations. When the robot does not conform to the rules of social interactions, an unwanted ‘hiccup’ in the conversation occurs where the interaction breaks down.

An important consideration in designing the robot’s behavioral system was that the robot can be given different movement animations to convey different emotions. For instance, the robot might shake visibly while picking up an object to show that it is uncertain about the correctness of the action. Also, the robot needs the capability to perform multiple actions simultaneously. For example, the robot should be able to look at the participant while pointing at an object of interest.

1.8 Experiment

The experiment was conducted in an Immersive Virtual Environment (IVE), in which participants interacted with a virtual robot to complete a cooperative task.

Since this was the first time the IVE at our disposal was used in a cooperative and social task, a framework of participant and robot interaction had to be constructed. Fundamentally, the framework had to provide two things: means for the participant to interact with objects in the virtual world, and a virtual social robot. The virtual robot needed to interact with objects in the virtual world, as well as with the participant.

A number of tasks were considered for the use in the experiment, before we settled on a fairly rigid nonverbal social decision task. A card was shown to the robot and the participant with a gesture on it, which they had to perform. Depending whether they made the same gesture or not, either the participant or the robot then had to pick up the card. Such a trial was repeated 60 times in the experiment.

The robot’s task-related performance behavior and movement style were varied as between-subject factors. Movement style was chosen as the aspect of ABS to be manipulated because it was relatively simple to implement. Moreover, the effect of behavioral and movement style on trust are not well understood, as opposed to appearance. The robot’s task performance was varied such that the robot performed well or badly on the task. The movement style could vary between smooth or shaky.

When the robot's movement style was shaky, it would sometimes perform its movements with a tremble.

The trust metric used in the experiment is an adaption of a questionnaire of human-computer trust, created by Madsen and Gregor (2000). They argue that trust is based on a number of different constructs, which are either cognitive based or affect based (see also Section 2.5). Cognitive-based constructs in the questionnaire are Perceived Reliability, Perceived Technical Competence and Perceived Understandability. Affect-based constructs are Faith and Personal Attachment.

The behavioral metrics are derived from the location and orientation of position-trackers worn by the participant in the IVE. Participants wore trackers on their head (on the head-mounted display), and their lower arms. This information was recorded during the duration of the experiment. With this information, it is possible to compute certain aspects of the participant's behavior. For instance, the amount of time the participant spends looking at the robot in a trial or the average speed he has when moving towards the robot can be calculated from the orientation and location of the head tracker.

For more information on the experimental procedure and method, see Chapter 2.

Overview

The structure of the rest of this thesis is as follows. The experimental setup and method is further explained in Chapter 2. The results of the experiment and statistical tests can be found in Chapter 3. These are discussed in Chapter 4. The thesis concludes with area's for further research in Chapter 5.

Chapter 2

Method

To measure the effect of different behavioral aspects of a social robot on trust, an experiment in a virtual reality environment was done. The participants of the Virtual Reality experiment performed a nonverbal social decision task in the form of a simple gesture recognition game with a virtual robot avatar.

2.1 Participants

Test subjects were students of Radboud University Nijmegen, which were awarded with a course credit or money for their participation.

2.2 Task

The task was situated in a virtual room, with a table in the center and the robot avatar on one side of the table. On the other side of the table, the participant's start position was marked with a traffic cone in the virtual room.

In each trial, a card appeared, standing vertically on the table in such a way that only one side could be seen by the participant. The task of the experiment was to decide whether the robot or the participant had to touch the card. The card showed one of two possible arm gestures. Participants were told that another gesture was shown to the robot. The participant and the robot each had to perform the gesture shown on their side of the card. When the robot and the participants made the same gesture, the participant had to touch the card on the table. When the gesture of the user and the robot were not the same, the robot had to pick up the card.

In reality, the image on the card had no meaning for the robot and its actions were fully scripted. Most of the time, the robot was instructed to make the 'correct' gesture for the trial, and the trial would end in success when the user also made the correct gesture. However, the robot could also be instructed to make a 'mistake' in recognizing the gesture on the card. In that case, the robot performed the wrong gesture and the card would be picked up by the wrong player. Whether a trial was successful or not was indicated by a 'happy' or 'sad' sound. The robot also portrayed joy or sadness depending on the outcome of the trial.

2.3 Conditions

Two factors of the robot’s behavior were manipulated in the experiment: the robot’s task performance and its movement style. If an agent makes few mistakes in a task, its task performance is said to be high or good. If an agent makes many mistakes in a task, its task performance is said to be low or bad. The movement style of an agent describes the way it moves to complete the task. The movement style can be used to express the level of certainty the agent has about his decisions. In our experiment, we let the robot move smoothly or shakily to convey this certainty.

The robot’s task performance was manipulated by letting the robot make more mistakes. Two levels of task performance were created: a robot with *good* performance, in which it performed 90% of all trials correctly, and a robot with *bad* performance on the task, in which the robot performed 70% of all trials correctly.

Two levels of movement style were created for the robot: the *smooth* style, in which the robot performed 100% of all trials with smooth motions, and a *shaky* style, in which the robot performed 70% of all trials with smooth motions and 30% of all trials with shaky motions.

With the smooth motion style, the robot performed his gesture and picked up the card fluently. In contrast, with the shaky motion style, the robot’s arms and body would shake when performing gestures and picking up the card. The shaky motions were created by using the smooth motions as a template, on top of which a sine wave with a frequency of 40 Hz to the rotation on all moving axes was added. The maximum amplitude of the wave was 1 degree over the roll and pitch axis, and 2 degrees over the yaw axis.

Each participant was presented with one level of each factor in the experiment.

2.4 Setup of the Immersive Virtual Environment

2.4.1 General Technical Information

The Immersive Virtual Environment (IVE) setup used in the RIVER lab (Radboud Immersive Virtual Environment Research lab) consists of a system that continuously (60 Hz) tracks the location and orientation of multiple trackers in an 8 x 6 meter room. One of these trackers is placed on top of a Head Mounted Display (HMD), displaying Open Competition MaGW/Open MaGW Program Application Form 2009 stereoscopic images to human participants. These images are rendered on-line based on the input of the tracking system by a high-end laptop worn on the back of a participant. Because the laptop is wirelessly connected to the tracking system, participants are able to move around freely within the virtual world. The location and orientation of the trackers can additionally serve as input for the virtual robot, present in the same virtual space as the human participants. Additional trackers can be placed on the participant to enable full body motion capturing, the data of which may serve as more complex input for the virtual robot.

2.4.2 Equipment

Two trackers were placed on the participant’s lower arms to track their location and orientation. This information was used to recognize the arm gestures made by the participant. In addition, the participant’s right hand was fitted with a 5DT Data Glove 5 MRI glove which registered the bending of the participant’s fingers. A virtual hand was linked to the tracker on the participant’s lower right arm and translated 15

centimeters to the right and 20 centimeters to the front of the tracker in the virtual world. The finger movements registered by the glove were used to bend the fingers of the virtual hand in the same way, so participants would see the hand in the virtual world as 'theirs'.

The use of a wrist tracker to link the hand to was considered in earlier stages of development of the pilot. However, this idea was discarded because the wrist tracker's small size caused it to be occluded by the participant's body for a significant amount of time. This caused the tracking system to lose track of the wrist tracker. The link to the lower arm tracker caused a loss in degrees of freedom of the hand because the movements of the participant's wrist could not be taken into account, but this proved not to be a serious issue in the selected task.

2.4.3 Gesture Recognition

The location data of the Head Mounted Display (HMD) and the two trackers on the participant's lower arms were used to recognize the arm gestures the participant made (Figure 2.1). Recognition of the 'Up' gesture, where the participant had to hold both his hands above his head, was triggered when at least one of the lower arm trackers was raised as high as 10 cm below the tracker on the HMD. The 'Wide' gesture, where the participant had to hold their arms to the side of their body, was triggered when the distance on the x-axis between both arm trackers was equal to or larger than 90 centimeters.

The use of these criteria resulted in a robust gesture recognition system which performed nearly optimally in the experimental setting. Occasional recognition errors were mostly caused by obscured trackers, but these happened very infrequently. Participants were instructed to practice the gestures before the real experiment began. Auditive feedback was given to the participant each time a gesture was recognized.

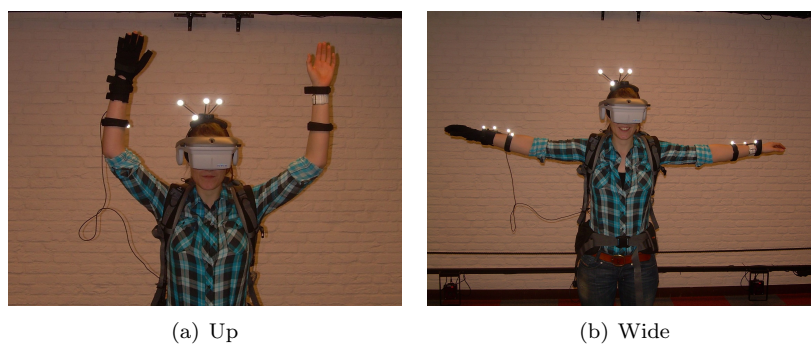


Figure 2.1: A participant demonstrating the 'arms up' (a) and 'arms wide' (b) gestures used in the experiment.

2.5 Metrics

Both trust and the behavior of the participant was measured in the experiment.

2.5.1 Trust

Trust was measured with a questionnaire which had to be filled out at the end of the experiment. The questionnaire presented to the participants consisted of three parts.

The first part of the questionnaire measured trust. The questions in this section measured five constructs of trust (Perceived Reliability, Perceived Technical Competence, Perceived Understandability, Faith, and Personal Attachment) and was adapted from Madsen and Gregor (2000). These constructs can be grouped into two categories of trust. Perceived Reliability, Perceived Technical Competence and Perceived Understandability measure cognition based trust, while Faith and Personal Attachment measure affect based trust. There were three items per construct, which the participant had to rate according to a 7-point Likert scale. The complete questionnaire (in Dutch) can be found in Appendix A.

The second and third part of the questionnaire consisted questions about agency and control questions to ensure the participant did not know what the experiment was about. These parts were not used in the analysis.

2.5.2 Behavior

The behavioral metrics are derived from the location and orientation of position-trackers worn by the participant in the IVE. Analyses were performed on four classes of behavioral metrics. Only trials where the participant removed the card from the table were used in the analyses. Reaction times and movement speeds of the participants could only be measured during these trials. Also, the participant was more active during these trials so more overt behavior should be present.

Four different metrics of behavior were derived from the IVE tracker data. These were:

Reaction times Although participants were not instructed to try to be as fast as possible, different reaction times were computed to see whether they would vary over the different factors of the experiment.

Looking behavior The looking behavior of each participant was derived from the orientation of the head tracker and the position of objects in the virtual environment. Although this does not capture exactly what a participant is looking at because eye movements were not measured, the orientation of the participant's head gives a rough estimate of what he is looking at. This metric is used to measure the neglect time.

Distances Interpersonal distances play an important role in human social interaction. That people tend to approach good friends more closely than acquaintances is well known. The distance between the participant and a trustworthy and untrustworthy robot might be similarly affected.

Movement speeds The approach and retreat speed of the participant can indicate the willingness to approach the robot.

Reaction time metrics were measured in seconds. The reaction times we tested for were:

Communication time The time from the appearance of the card to the time where the robot and participant both performed their gesture.

Card time The total time the card is on the table.

Time to gesture The time from the appearance of the card to the moment where the participant makes his or her gesture.

Pick card time The time it takes from the moment the participant makes his gesture to the time he picks up the card.

Time to closest approach The time it takes the participant to move from his start position to his closest approach with the robot.

Time to start position The time it takes for the participant to move back to his start position after the card was removed from the table.

Metrics for the looking behavior were:

percent looking at robot The percentage of time the participant looks at the robot during the duration of the trial.

number of times looking at robot The average time a participant starts looking at the robot. This measures the number of times the participant looks at the robot during a trial. When the value is low, the participant looks at the robot for a long duration, while the look duration is shorter when this value is high. This metric is divided by the duration of a trial to correct for the varying duration of the trials.

It should be noted that the metrics for the looking behavior were not very precise because only the orientation of the participant's head was known. The assumption is made that the head's orientation roughly corresponds with the gaze direction of the participant.

Metrics for distances were measured in meters. The following metrics were used:

Starting distance The distance from participant's start point to robot.

Minimum distance The distance from closest approach of participant to robot.

Walking distance The distance between start position and closest approach.

Metrics of movement speeds were:

speed to closest approach The average walking speed of the participant towards the closest approach of the robot.

speed back to start position The average walking speed of the participant back to the start position after the card has been picked up.

Both metrics are reported in meters per second (m/s). The moment the participant starts walking is defined as the moment when he or she leaves the area described by a circle with a radius of 0.5 m centered on the participant's starting position.

2.6 Experiment Procedure

Participants were given a short tour of the lab after filling in a pre-screen form. After all equipment was fitted to the participant, they were given a short time to get acquainted with the virtual world and were made comfortable moving around in the environment. All participants were then instructed how to make the gestures used in the task and how to pick up the card with their right hand.

Following this, a short practice block of eight trials was started to practice the task. The practice block was identical for all participants. The experimenter was present in the room with the participant, so he could give additional instructions or answer any questions the participant might have. In one of the trials in the training

block the robot made a mistake. Participants were asked if they understood why the trial ended in failure and were explained that the robot was the one who made the mistake in this instance.

After the practice block, the experiment was started. The experimenter left the experiment room, and monitored the participant's progress from an adjacent control room. A new robot, which had a slightly different color scheme than the practice robot was shown to the participants. This was done so the participant would not have any expectations about the robot's performance, which may have been present after the practice block.

The experiment consisted of 60 trials per participant, divided in 6 blocks of equal length. Between blocks, there was a short break, during which the experimenter could make adjustments to the trackers if this was necessary. The participant could also sit down for a moment if he or she felt uncomfortable with the backpack. Most of the time, however, adjustments or breaks were not needed and the experimenter would start the next block right after the previous block was completed.

In the first trial block, the robot never made mistakes meaning its task performance was optimal. The robot's movement style was the same in all blocks according to the selected confidence style. This was done so the participant could bond with and learn to trust the robot. In the next five blocks, the robot made mistakes dependent on the selected level of task performance (see Section 2.3).

After the experiment was finished, participants were taken to an other room and had to fill in the questionnaire to measure their explicit trust in the robot.

Chapter 3

Results

Below, the results of the experiment are presented. Information about the participants and exclusion criteria are presented in Section 3.1. Information about the statistical tests are found in Section 3.2. The effects of task performance and movement style on trust are presented in Section 3.3, and their effects on behavior can be found in Section 3.4. Correlation and mediation results of trust and behavior are located in Section 3.5.

3.1 Participants

74 people were tested in the experiment. 52 participants were Dutch, 20 German and 1 Iranian and 1 Romanian. During the experiment it became clear that foreign students did not understand the trust questionnaire, which was in Dutch, well enough. Therefore, it was decided only to use the data from Dutch participants for analysis. Of the 52 Dutch participants, 3 more were excluded. Two participants told the experimenter they did not understand the task after they completed the experiment, and one told the experimenter he realized what the experiment was about and had answered the questionnaire accordingly.

3.2 Test Information

2 (task performance: bad vs. good) x 2 (movement style: shaky vs. smooth) between-subjects ANOVA on various trust metrics were conducted. When post-hoc tests required pairwise comparisons, independent sample *t*-tests were performed.

3.3 Explicit Trust

The internal reliability of the items was measured with Cronbach's alpha. While Perceived reliability and Perceived Technical Competence can be classified as reliable (Perceived Reliability: .761, Perceived Technical Competence: .606), the other constructs had low internal consistency (Perceived Understandability: .534, Faith: .246, Personal Attachment: .412).

The average of all items was taken and this measure was used as the total trust score of a participant. Consistency of total trust is good (Cronbach's alpha: .752). The average scores of items in the cognition and affect based trust categories were used as the trust score for their respective categories. Combined reliability of cognition

based trust was high (Cronbach’s alpha: .808), while the reliability of affect based trust was low (Cronbach’s alpha: .518).

3.3.1 Analysis of total trust

The 2 (task performance: bad vs. good) x 2 (movement style: shaky vs. smooth) between-subjects ANOVA on the total trust score revealed a significant main effect of task performance, $F(1, 45) = 4.86$, $p < .05$, with an effect size of $\eta_p^2 = .10$. On average, a robot with a good task performance scored higher ($M = 4.98$, $SE = 0.63$) than a robot with bad task performance ($M = 4.70$, $SE = 0.67$). Also, a significant task performance x movement style interaction effect was found, $F(1, 45) = 6.042$, $p < .005$, with an effect size of $\eta_p^2 = .20$. The directions of these effects are described in Table 3.1.

The result of the analysis of total trust show both a performance and a fit effect. On average, a robot with good task performance is trusted more than a robot with bad task performance. Additionally, a robot which task performance matches its movement style is trusted more than a robot which task performance and movement style are not consistent.

Total Trust	Task performance	
	bad	good
Movement style		
smooth	$M = 4.29^{\star\dagger}$ $SE = 0.67$	$M = 5.24^{\star\star}$ $SE = 0.49$
shaky	$M = 4.97^{\dagger}$ $SE = 0.51$	$M = 4.28^{\star}$ $SE = 0.67$

Table 3.1: The means (M) and standard errors (SE) of Total trust in each condition. Means denoted with a \star , \ast or \dagger differ significantly from each other. \ast is marginally significant.

3.3.2 Differences between cognition and affect based trust

Madsen and Gregor (2000) make a distinction between cognitive and affective trust. They argue that the constructs Perceived Reliability, Perceived Technical Competence and Perceived Understandability relate to a cognitive approach towards trust, while Faith and Personal Attachment are more related to affective based trust. To test for differences between cognitive and affective trust, 2 (task performance: bad vs. good) x 2 (movement style: shaky vs. smooth) between-subjects ANOVAs were performed on the averaged scores of all items related to the different concepts of trust.

In the case of cognitive trust, the test showed a significant effect of task performance, $F(1, 45) = 6.572$, $p < 0.025$, with an effect size of $\eta_p^2 = .13$. On average, a good Performance of the robot leads to higher cognitive trust ($M = 5.28$, $SE = 0.70$) than a bad task performance ($M = 4.82$, $SE = 0.83$). Also, a significant task performance x movement style interaction effect was found, $F(1, 45) = 6.70$, $p < 0.05$ with an effect size of $\eta_p^2 = .13$. Post hoc tests are described in Table 3.2.

When looking at affect based trust, only a significant task performance x movement style interaction effect was found, $F(1, 45) = 9.55$, $p < .05$, effect size $\eta_p^2 = .18$. The results of the post hoc tests are summarized in Table 3.3.

Cognition based Trust	Task performance	
	bad	good
Movement style		
smooth	$M = 4.51^{\star\dagger}$ $SE = 0.83$	$M = 5.60^{\star\star}$ $SE = 0.62$
shaky	$M = 5.02^{\dagger}$ $SE = 0.53$	$M = 5.01^{\star}$ $SE = 0.91$

Table 3.2: The means (M) and standard errors (SE) of Cognition based trust in each condition. Means denoted with a \star , \ast or \dagger differ significantly from each other. \ast and \dagger are marginally significant.

Affect based Trust	Task performance	
	bad	good
Movement style		
smooth	$M = 3.95^{\star\dagger}$ $SE = 0.66$	$M = 4.70^{\star}$ $SE = 0.69$
shaky	$M = 4.89^{\star\dagger}$ $SE = 0.53$	$M = 4.40^{\star}$ $SE = 0.91$

Table 3.3: The means (M) and standard errors (SE) of Cognition based trust in each condition. Means denoted with a \star , \ast or \dagger differ significantly from each other. \ast is marginally significant.

These results show that a fit effect is present in both cognition based and affect based trust scores, although the effect size is somewhat smaller $\eta_p^2 = .13$ for cognition based trust than for affect based trust ($\eta_p^2 = .18$). Only cognition based trust showed a significant effect of task performance. All trust variables are compared in Figure 3.1.

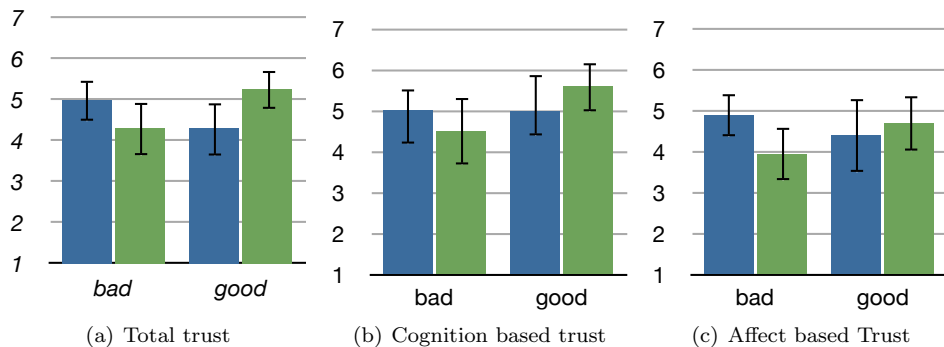


Figure 3.1: Graphs of trust metrics. Blue bars represent a shaky movement style, green bars a smooth movement style. Both the total trust (a) and cognition based trust (b) show a performance effect, while this is absent in affect based trust (c). All trust variables show a similar task performance \times movement style interaction pattern.

3.4 Behavioral metrics

For all reported variables, their log values were taken to reduce the skewness. These variables were then standardized to Z-scores. All trials which took longer than 3 standard deviations of the mean trial time were excluded from the analyses. Of all valid measurements, the mean for each subject was calculated. Analyses were done on the standardized Z-scores of the logs of each reported variable. However, the reported means are translated back to the normal value because it is easier for interpretation.

3.4.1 Reaction times

Analysis showed that there was a marginally significant Performance effect for Pick card time ($F(1, 45) = 3.09, p < .10, \eta_p^2 = .06$). On average, the time for a participant to pick up the card after he or she made the gesture was shorter when the robot's task performance was good ($M = 4.21, SE = 0.47$) than when the robot's task performance was bad ($M = 4.55, SE = 0.96$).

There was also a significant Performance \times Style interaction effect for the time to gesture metric ($F(1, 45) = 4.40, p < .05, \eta_p^2 = .09$). Post-hoc test results are summarized in Table 3.4.

Time to gesture Movement style	Task performance	
	bad	good
smooth	$M = 1.37^{\star\dagger}$ $SE = 0.16$	$M = 1.47^{\star}$ $SE = 0.36$
shaky	$M = 1.55^{\dagger}$ $SE = 0.21$	$M = 1.50$ $SE = 0.19$

Table 3.4: The means (M) and standard errors (SE) of Time to gesture in each condition. Means denoted with a \star or \dagger differ significantly from each other. \star and \dagger are marginally significant.

Analysis of time to closest approach revealed a significant Performance \times Style interaction effect, $F(1, 45) = 5.07, p < .05, \eta_p^2 = .10$. Results of the post hoc tests can be found in Table 3.5. On average, participants were faster in reaching their closest approach with the robot when its movement style was smooth than when its movement style was shaky when task performance was good. Also, a marginal significant effect of task performance was found when the robot's movement style was shaky. Participants were faster in reaching their minimum distance when the robot's task performance was good compared to bad task performance when the robot's movement style was shaky.

A significant Performance \times Style interaction effect was also found for the time to start position, $F(1, 45) = 8.97, p < .05, \eta_p^2 = .16$. Results of the post-hoc tests can be seen in Table 3.6.

3.4.2 Looking Behavior

There was a marginally significant effect of movement style on the number of times a participant looked at the robot, $F(1, 45) = 2.98, p < .10, \eta_p^2 = .07$. On average, participants started looking more often at the robot when its movement style was shaky ($M = 0.33, SE = 0.13$) than when its movement style was smooth ($M = 0.40, SE = 0.17$).

Time to closest approach	Task performance	
Movement style	bad	good
smooth	$M = 1.63^{*\dagger}$ $SE = 0.42$	$M = 1.80^*$ $SE = 0.19$
shaky	$M = 1.81^\dagger$ $SE = 0.31$	$M = 1.80$ $SE = 0.23$

Table 3.5: The means (M) and standard errors (SE) of time to closest approach in each condition. Means denoted with a $*$ or \dagger differ significantly from each other. $*$ and \dagger are marginally significant.

Time to start position	Task performance	
Movement style	bad	good
smooth	$M = 2.03^{*\dagger}$ $SE = 0.42$	$M = 2.38^{**}$ $SE = 0.47$
shaky	$M = 2.44^{\ddagger\dagger}$ $SE = 0.31$	$M = 2.07^{\ddagger*}$ $SE = 0.28$

Table 3.6: The means (M) and standard errors (SE) of Time to start position in each condition. Means denoted with a $*$, $*$, \dagger or \ddagger differ significantly from each other. $*$ and \dagger are marginally significant.

3.4.3 Distances

Analysis showed a marginally significant effect of style on minimum distance, $F(1, 45) = 3.51$, $p < .10$, $\eta_p^2 = .07$. On average, participants approached a smooth moving robot more closely ($M = 2.24$, $SE = 0.13$) than a shaky robot ($M = 2.30$, $SE = 0.09$).

A significant Performance \times Style interaction effect was found for starting distance, $F(1, 45) = 2.54$, $p < 0.005$, $\eta_p^2 = .09$. Results of the post-hoc tests can be found in Table 3.8.

Starting distance	Task performance	
Movement style	bad	good
smooth	$M = 3.46$ $SE = 0.24$	$M = 3.55^*$ $SE = 0.19$
shaky	$M = 3.62^\dagger$ $SE = 0.20$	$M = 3.46^{\dagger*}$ $SE = 0.23$

Table 3.7: The means (M) and standard errors (SE) of Starting distance in each condition. Means denoted with a $*$ or \dagger differ significantly from each other. $*$ is marginally significant.

For walking distance, a significant Performance \times Style interaction effect was found as well $F(1, 45) = 6.61$, $p < .025$, $\eta_p^2 = .13$. Results of the post-hoc tests are summarized in Table 3.8.

Walking distance	Task performance	
	bad	good
smooth	$M = 1.19$ $SE = 0.27$	$M = 1.34^*$ $SE = 0.43$
shaky	$M = 1.33^\dagger$ $SE = 0.21$	$M = 1.14^{\dagger*}$ $SE = 0.34$

Table 3.8: The means (M) and standard errors (SE) of Walking distance in each condition. Means denoted with a \star or \dagger differ significantly from each other. \dagger is marginally significant.

3.4.4 Movement speeds

Performance \times Style interaction effects were found for both speed metrics. In the case of the speed to closest approach, the effect was marginally significant, $F(1, 45) = 3.75$, $p < .10$, $\eta_p^2 = .07$. Results of the post-hoc tests can be seen in Table 3.9.

Speed to closest approach	Task performance	
	bad	good
smooth	$M = 0.45$ $SE = 0.11$	$M = 0.49^*$ $SE = 0.09$
shaky	$M = 0.47^\dagger$ $SE = 0.08$	$M = 0.42^{\dagger*}$ $SE = 0.10$

Table 3.9: The means (M) and standard errors (SE) of Speed to closest approach in each condition. Means denoted with a \star or \dagger differ significantly from each other. Both effects are marginally significant.

The Performance \times Style interaction effect for speed to start position was significant, $F(1, 45) = 4.94$, $p < .05$, $\eta_p^2 = .10$. Results of the post-hoc tests can be found in Table 3.10.

Speed to start position	Task performance	
	bad	good
smooth	$M = 0.09$ $SE = 0.14$	$M = 0.17^*$ $SE = 0.10$
shaky	$M = 0.17^\dagger$ $SE = 0.09$	$M = 0.06^{\dagger*}$ $SE = 0.10$

Table 3.10: The means (M) and standard errors (SE) of Speed to start position in each condition. Means denoted with a \star or \dagger differ significantly from each other. Both effects are marginally significant.

3.4.5 Summary

Several behavioral metrics showed significant or marginally significant effects of movement style or Performance \times Style interaction. The number of times looking at robot

and minimum distance metrics had a significant movement style effect. Time to gesture, time to closest approach, time to start area, start distance, walking distance, speed to closest approach and speed to start position all had a significant Performance \times Style interaction. Interestingly, the interaction patterns were largely similar as well. Time to pick card was the only metric that showed a marginally significant effect of performance.

3.5 Correlation and Mediation of Trust on behavior

The question still remains whether any of the behavioral effects described above correlate with the trust metrics from the questionnaire (Section 3.3). Therefore, correlation and mediation analyses were performed on all (marginally) significant behavioral and trust metrics.

3.5.1 Correlations

Correlation between the behavioral variables was computed using Pearson's coefficient. The resulting correlation matrix can be found in Table 3.11. The correlations between trust and behavioral variables can be found in Table 3.12.

	TG	TCA	TS	TLR	MD	SD	WD	SCA	SS
TPC	.302*	.331*	.394**	.184	.024	.377**	.309*	.146	.238
TG		.294*	.428**	.203	.549**	.198	.399**	.319*	.340*
TCA			.871**	.115	-.268	.776**	.805**	.299*	.687**
TS				.109	.436**	.687**	.808**	.436**	.691**
TLR					-.120	.281	.279	.303*	.207
MD						-.027	-.457**	-.445**	-.432**
SD							.896**	.656**	.813**
WD								.799**	.938**
SCA									.855**

Table 3.11: Correlation matrix of significant behavioral metrics. * $p < .05$, ** $p < .01$. TPC = Time to pick card, TC = Time to gesture, TCA = Time to closest approach, TS = time to start, TLR = times looking at robot, MD = minimum distance, SD = start distance, WD = walking distance, SCA = speed to closest approach, SS = speed to start.

The correlation matrices show that strong correlations exist, especially between behavioral metrics which are dependent on distances. Strong correlations also exist between the behavioral and trust metrics, which suggest that they are similarly manipulated by task performance and movement style.

3.5.2 Mediation

Mediation analyses were performed to see whether the effects of task performance and movement style on behavior take an indirect path via trust. A mediation analysis is done by constructing two linear models, one without and one with the mediator variable. A prerequisite is that both the dependent variable and the mediator variable are manipulated by the independent variables. In the first model, the direct effect of the independent variables on the dependent variable is tested. In the second model, an intermediate variable is introduced. If the direct effect is no longer significant and the indirect effect is shown to exist, the intermediate variable mediates the effect of the independent variables on the dependent variable (see also Figure 3.2).

Behavioral variable	Cognition based trust	Affect based trust	Total trust
Time to pick card	-.008	.170	.072
Time to gesture	.169	.249 [†]	.237
Time to closest approach	.395**	.386**	.464**
Time to start	.395**	.425**	.482**
Times looking at robot	.143	.124	.160
Minimum distance	-.289*	-.050	-.233
Starting distance	.283*	.309*	.347*
Walking distance	.408**	.304**	.347*
Speed to closest approach	.304*	.113	.272 [†]
Speed to start	.462**	.202	.429**

Table 3.12: Correlations between trust and behavioral metrics. [†] $p < .10$, * $p < .05$, ** $p < .01$

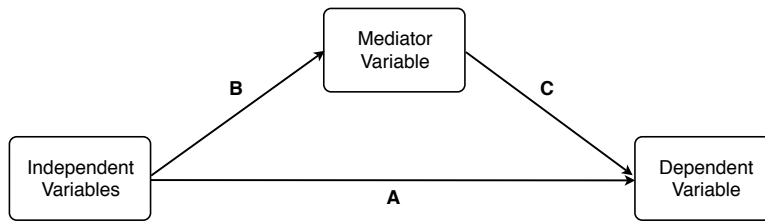


Figure 3.2: A general model of mediation. A is the direct effect of the independent variables on an independent variable. Full mediation exists when it can be shown effects B and C are significant, while A is not.

In Sections 3.3.1 and 3.3.2 it was established that there are main effects of the dependent variables task performance and movement style on all trust variables. Several behavioral variables were shown to be manipulated by task performance and movement style as well (Section 3.4). Therefore, we can use the trust variables in a mediation analysis to test if they mediate the effect of task performance and movement style on the significant behavioral variables.

On all behavioral variables on which the task performance and movement style had a significant or marginally significant effect, two regression analyses were performed to test for mediation. First, the direct model with only the effects of task performance, movement style and task performance \times movement style interaction was done. Second, the trust variable was entered into the model to see whether it is a mediator for the behavioral variable. In Table 3.13, the results of these analyses are summarized. The complete multiple regression models for each mediated variable can be found in Appendix B.

Behavioral metric	Affect based trust	Cognition based trust	Total trust
Time to closest approach	✓	✓	✓
Time to start position	✓	✓	✓
Starting distance		✓	✓
Walking distance		✓	✓
Speed to closest approach		✓	
Speed to start position		✓	✓

Table 3.13: A summary of the mediation analyses. A checkmark (✓) means the behavioral metric is mediated by the corresponding trust variable.

Switching the mediator with the independent variables in the analysis showed the behavioral metrics also have predictive properties for the total trust score. However, the regression weight of task performance \times movement style interaction effect stays significant. This means that behavioral variables do not mediate the effects of task performance and movement style on trust.

Chapter 4

Discussion

In this chapter, the results from the experiment and their implications are discussed. First, a brief summary of the results is presented in Section 4.1. Then, the implications of the results on trust are discussed in Section 4.2. The role of trust on behavior is discussed in Section 4.3. This is then followed by the general conclusions (Section 4.4).

4.1 Summary of results

In the previous chapter, significant interaction effects were found for all trust variables and several behavioral variables.

For the total trust and cognitive trust scales, the robot is trusted more when its task performance was good. Also, the robot is trusted more when its task performance and its movement style were consistent. A smoothly moving, well performing robot and a shaky moving, badly performing robot are trusted more than a shaky robot with good task performance and a smooth robot with a bad task performance.

On a behavioral level similar patterns are found. When the robot's task performance and movement style are consistent, participants are a little slower, stay slightly further away from the robot and have higher movement speeds. When task performance and movement style was inconsistent, participants are somewhat faster, are somewhat closer to the robot and have slightly slower movement speeds.

Roughly two thirds of all post-hoc tests which reported significant differences were only marginally significant. This can be attributed to the fact that only the data of Dutch participants was used. Increasing the sample size by testing more participants will probably alleviate this problem.

4.2 Effects on Trust

Effects of explicit trust found in the analysis of the questionnaire. Two main effects were found the total trust score. First, there is an effect of task performance. When a robot has a good task performance, it is generally regarded as more trustworthy than when it has a bad task performance. Second, there is a main interaction effect of task performance and movement style on trust. This is a fit effect, meaning that when a robot's task performance and movement style are consistent (bad performance combined with shaky movements or good performance combined with smooth movements), the robot is trusted more than when its performance and style are inconsistent

(bad performance combined with smooth movements or good performance combined with shaky movements).

When the total trust score is split up between its cognition based and affect based trust constructs, it can be established that there is only an effect of task performance on cognitive trust. This effect is not present in affect based trust. Also, none of the behavioral variables which trust mediates were shown to have a performance effect. This suggests that the performance effect is based on higher-level reasoning about the robot's actions.

These results show that while it is important to perform well on a task in order to be trusted, movement style also play an important role when perceiving trustworthiness. Our social robot is trusted more when its behavioral style conveys a level of certainty consistent with its task performance. This suggests that while it is best to have a well performing social robot with a smooth and confident movement style, it is also very important for a social robot to communicate uncertainty when it cannot perform well on a task.

This effect of consistent appearance and performance has been described in the literature before. While this article describes differences in attitude based on appearance, the pattern is similar to the one found in this research. Richman (1978) let teachers score the intelligence level of children with a cleft palate. The children were placed in two groups – one group of children had a relatively normal facial appearance, while the second group consisted of children with more noticeable facial disfigurement. The groups did not differ intellectually or behaviorally. The teachers' ratings of the children's intelligence were compared to their objective IQ scores. Children with high IQ scores and normal appearance were rated higher than smart children with noticeable facial disfigurement. However, when children with low IQ scores were rated, the group with more facial disfigurement was rated higher by the teachers than children with a normal appearance.

The behavioral style of a trustee is used by the trustor as an indicator of its task competence. When they are aligned, the trustor is able to make a good assessment of the capabilities of the trustee, which results in higher trustworthiness. Being able to make a correct assessment of the trustee's task performance on the basis of its behavioral style is a trustworthy quality. When the behavioral style of the trustee are not aligned, the trustor cannot make a reliable assessment of the trustee's capabilities. This lowers the trustor's perceived trustworthiness of the trustor.

4.3 Role of Trust on Behavior

Tests established mediation of explicit trust metrics on behavior of the participants. Behavior mediated by trust has mainly to do with the distance the participant keeps to the robot and its movement speeds. When the participant places more trust in the robot, he or she stays further away from the robot and moves faster than when his or her trust in the robot is lower. This means that when the trust score of a participant is known, a prediction can be made about his or her distances and movement speeds to and from the robot. The opposite is also true; from behavioral metrics, an estimate of the trust level of the participant can be made.

Mediation of trust on behavior does not imply that there is a causal relationship of trust on behavior (e.g., manipulations of task performance and movement style cause changes in trust which causes change in behavior). To test for causality, trust has to be directly manipulated in an experiment. However, analyses show that behavioral metrics do not mediate trust. While this does not qualify as evidence for causality, it does suggest that the effect of trust accounts for the changes in behavior and not the

other way around.

Interestingly, affect based trust mediated only two behavioral variables, which is much less than total trust (5) and cognition based trust (6). Since no effect of task performance was found for affect based trust and most behavioral variables, one could expect it would mediate better than cognition based and total trust. However, this was not the case. One of the reasons for this is that for affective trust, there was no difference in trust between movement styles when the robot had good task performance. There is a significant difference of the participants' trust score in the other trust variables as well as for most behavioral variables. Because of this difference in the interaction pattern, affect based trust mediates fewer behavioral variables than cognition based and total trust.

The participant's score on explicit trust variables are a good predictor for behavioral metrics. A very important implication of this result is that a social robot can measure these behaviors of the participant in real time and use this information to determine how well it is being trusted by the participant. With the reverse use of this information, the robot can assess whether it is being trusted well enough for its current task performance. When this is not the case and the participant is under- or overtrusting the robot, it can change its behavioral style so that the right amount of trustworthiness is again being perceived by the participant. This will then result in recalibration of participant's trust in the robot, thus leading to better calibrated trust.

4.4 Conclusions

In conclusion, this research has led to two important findings. First, it is important to align a trustee's movement style to its task performance. A bad movement style should serve as an indication that the trustee's task performance is low.

Second, the perceived trustworthiness of the trustee by the trustor is expressed in the behavior of the trustor. This can be an important tool in Social Robot research, because a social robot can use this information to calibrate its trustworthiness.

Chapter 5

Further Research

Hopefully, the results of the experiment in this thesis will be used in future work. As with all research, the results of this study have led to several new ideas which can be explored in future work.

5.1 Aspects of Appearance and Behavioral Style

The experiment in this thesis only manipulated the movement style of the robot to create a difference in behavioral style. There were several reasons for this. First, it is not possible to alter the physical appearance of a robot on the spot. While it is interesting to study the effect of differently shaped robots on the trustworthiness, the appearance of a single robot cannot be changed. In a real world scenario, the only thing that can be altered is the robot's behavioral style.

Second, we chose to manipulate the robot's movement style because it was the easiest aspect of behavioral style to vary. Other aspects of behavioral style, such as posture or turn taking behavior are harder to manipulate because they require more programming to achieve and knowledge to implement convincingly.

In future research, different aspects of behavioral style should be manipulated to build a more solid foundation for the hypothesis that appearance and behavioral style have an effect on the trustworthiness of the robot, and not just a shaky or smooth movement style.

5.2 Change of trust over time

In the current experiment, trust was measured after the entire task was completed. Moreover, behavioral variables are the result of averaging over all trials. Because trust is the result of observed task performance and behavior, the trust attitude of the participant is likely to change over the duration of the experiment as the participant observes more and more task performance. It would be interesting to see whether and how the participant's trust attitude changes over time.

It would be cumbersome to present the participant with a trust questionnaire after the completion of a block of trials. Additionally, it could give the participant an idea of what the experiment is about which could result in different behavior and attitude. However, since there is a relationship between trust and behavior, it is possible to derive the participant's trust attitude from his behavior. The behavior of a participant is recorded unobtrusively in a IVE, as was done in this study. The

behavioral variables can be compared after every block. The data from the current study is not suited for this, because trials were randomized over the entire duration of the experiment. The results will be best when every block of trials contains the same amount of manipulations for movement style and task performance.

5.3 Looking behavior and neglect time

The average looking time of the participant to the robot did not differ significantly over conditions. Looking time and neglect time are important metrics for trust in human robot interaction, and was expected to play a role in this experiment as well. Neglect time is used in human-robot interaction studies because the metric is not as task specific as, for instance, reaction times (Steinfeld et al., 2006). A trustworthy robot was expected to be neglected more than an untrustworthy robot, but this hypothesis is not supported by our results. This can be caused by two things. The setup of the experiment made looking behavior an unsuitable metric, or looking behavior and neglect time are unsuitable metrics for a cooperative task.

The experimental setup of this study could have made looking behavior and neglect time an unsuitable metric. In the experiment, participants always faced the robot. The robot was always in full view, in the center of the participant's vision. In addition, measuring gaze direction on the basis of the head tracker was not precise because the Head Mounted Display only registered the orientation of the head and not the direction of the participant's eyes. Furthermore, there was little else for the participant to focus their attention to do when the experiment was taking place except for the robot and the card.

In order to get better results with the looking behavior and neglect time, several things can be done. First of all, it is possible to build an eye tracker into the Head Mounted Display of an IVE (Duchowski et al., 2000). This makes it possible to record the participant's looking behavior more precisely. Second, changes to the experimental setting can be made so the participant does not face the robot all the time. When the participant has to make an effort to get the robot in view, the looking behavior will likely change. Third, participants can be given a secondary task in the experiment, such as completing simple math problems. The neglect time of the robot can then be expressed as the participant's score on the secondary task.

The other possibility is that neglect time is not a good metric for a cooperative task. Neglect time measures the capacity of a system to complete a task without the intervention of a human operator. Although this is an important factor when the system needs to be independent, it might not be as useful when the task requires full cooperation between the robot and its operator, as is the case in this experiment.

5.4 Realism

While more lifelike scenarios were considered as a task for the experiment, such as collaborating in packing a suitcase with the necessary items for a holiday, it was ultimately decided not to use this approach. Such a task might be more realistic in nature, but it would also delegate a lot of control to the participant. Since we were more interested in the direct effects of the robot's task performance and behavioral style, a more conventional task in which the control was less in the hands of the participant was chosen.

Now that the basic effects of task performance and movement style are known, more realism can be added to future experiments. This is especially important because

the ultimate goal of the Trust Me research project is to create a social robot that can function in a real world environment.

References

- Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, *13*(2), 103–124.
- Breazeal, C. L. (2002). *Designing sociable robots*. Boston, Massachusetts: MIT Press.
- Breazeal, C. L. (2003). Toward sociable robots. *Robots and Autonomous Systems*, *42*, 167–175.
- Clifford, M. M., & Walster, E. (1973). The effect of physical attractiveness on teacher expectations. *Sociology of Education*, *46*(2), 248–258. Available from <http://www.jstor.org/stable/2112099> (ArticleType: primary_article / Full publication date: Spring, 1973 / Copyright 1973 American Sociological Association)
- Dotsch, R., & Wigboldus, D. H. J. (2008). Virtual prejudice. *Journal of Experimental Social Psychology*, *44*, 1194–1198.
- Duchowski, A. T., Shivashankaraiah, V., Rawls, T., Gramopadhye, A. K., Melloy, B. J., & Kanki, B. (2000). Binocular eye tracking in virtual reality for inspection training. In *Proceedings of the 2000 symposium on eye tracking research & applications* (pp. 89–96). Palm Beach Gardens, Florida, United States: ACM. Available from <http://portal.acm.org/citation.cfm?id=355031>
- Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robots and Autonomous Systems*, *42*, 143–166.
- Forlizzi, J. (2007). How robotic products become social products: an ethnographic study of cleaning in the home. In *Proceedings of the ACM/IEEE international conference on human-robot interaction* (pp. 129–136). Arlington, Virginia, USA: ACM. Available from <http://portal.acm.org/citation.cfm?id=1228734>
- Goodrich, M. A., & Schultz, A. C. (2007). Human-robot interaction: A survey. *Human-Computer Interaction*, *1*(3), 203–275.
- Haselager, W., Wigboldus, D., & Dotsch, R. (2009). *Trust me! how robots earn human trust*. Unpublished NWO Program Application. Nijmegen, Netherlands.
- Iwata, H., & Sugano, S. (2009). Design of human symbiotic robot TWENDY-ONE. In *Proceedings of the 2009 IEEE international conference on robotics and automation* (pp. 3294–3300). Kobe, Japan: IEEE Press. Available from <http://portal.acm.org/citation.cfm?id=1703981>
- Lee, J., & See, K. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*, 50–80.
- Madsen, M., & Gregor, S. (2000). Measuring human-computer trust. In *11th australasian conference on information systems* (Vol. 53).
- Mori, M. (1970). Bukimi no tani [The uncanny valley] (K. F. MacDorman & T. Minato, Trans.). *Energy*, *7*(4), 33–35. Available from <http://www.androidscience.com/theuncannyvalley/proceedings2005/uncannyvalley.html>

- Richman, L. C. (1978). The effects of facial disfigurement on teachers' perception of ability in cleft palate children. *Cleft Palate J*, *15*, 155-160.
- Scheef, M., Pinto, J., Rahardja, K., Snibbe, S., & Tow, R. (2002, may). Experiences with Sparky, a social robot. In K. Dautenhahn (Ed.), *Socially intelligent robots: creating relationships with computers and robots* (pp. 173-180). Dordrecht, the Netherlands: Kluwer Academic Publishers Group.
- Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., et al. (2006). Common metrics for human-robot interaction. In *Hri '06: Proceedings of the 1st acm sigchi/sigart conference on human-robot interaction* (pp. 33-40). New York, NY, USA: ACM.
- Yuki, M., Maddux, W. W., Brewer, M. B., & Takemura, K. (2005). Cross-Cultural differences in relationship- and Group-Based trust. *Pers Soc Psychol Bull*, *31*(1), 48-62. Available from <http://psp.sagepub.com/cgi/content/abstract/31/1/48>

Appendices

Appendix A

Questionnaire

Below is the questionnaire with the trust constructs as presented to the participants of the experiment. The questionnaire was adapted from Madsen and Gregor (2000) and was translated to Dutch.

Vragenlijst

Geef van elk van onderstaande stellingen aan in hoeverre jij vindt dat hij van toepassing is. Doe dit door telkens n van de cijfers te omcirkelen.

1. De robot gedraagt zich betrouwbaar.

helemaal niet 1 2 3 4 5 6 7 heel erg

2. De robot interpreteert de situatie telkens consistent.

helemaal niet 1 2 3 4 5 6 7 heel erg

3. De robot geeft me de indruk te weten op welke manier hij zich moet gedragen bij dit probleem.

helemaal niet 1 2 3 4 5 6 7 heel erg

4. De robot gebruikt mijn input goed in zijn eigen functioneren.

helemaal niet 1 2 3 4 5 6 7 heel erg

5. Ik begrijp de robot in zijn gedrag zodat ik de volgende keer goed kan samenwerken met de robot.

helemaal niet 1 2 3 4 5 6 7 heel erg

6. Het is gemakkelijk om te volgen wat de robot doet.

helemaal niet 1 2 3 4 5 6 7 heel erg

7. Wanneer ik onzeker was over wat er van mij werd verwacht kon ik op het gedrag van de robot vertrouwen om te zien wat ik moest doen.

helemaal niet 1 2 3 4 5 6 7 heel erg

8. Zelfs ook al geeft de robot me geen reden om te verwachten dat het de taak goed zal uitvoeren, toch heb ik het vertrouwen dat de robot het goed doet.

helemaal niet 1 2 3 4 5 6 7 heel erg

9. Ik voelde een bepaalde soort van verbondenheid met de robot in de taak van zojuist.

helemaal niet 1 2 3 4 5 6 7 heel erg

10. Ik vind het leuk om samen met de robot tot een beslissing te komen.

helemaal niet 1 2 3 4 5 6 7 heel erg

11. De robot had de intentie om goed te presteren op de taak die we zojuist hebben uitgevoerd.

helemaal niet 1 2 3 4 5 6 7 heel erg

12. Ik had het gevoel dat de gezamenlijke prestaties op de taak afhankelijk waren van de handelingen van de robot.

helemaal niet 1 2 3 4 5 6 7 heel erg

13. Ik had het idee dat de robot betrokken was bij het uitvoeren van de taak.

helemaal niet 1 2 3 4 5 6 7 heel erg

14. Ik had de intentie om goed te presteren op de taak die we zojuist hebben uitgevoerd.

helemaal niet 1 2 3 4 5 6 7 heel erg

15. Ik had het gevoel dat de prestaties op de taak afhankelijk waren van mijn handelingen.

helemaal niet 1 2 3 4 5 6 7 heel erg

Appendix B

Mediator Regression Models

Below the linear models used in the mediation analyses of the behavioral metrics. The procedure of the process is described in Section 3.5.2.

B.1 Mediation of Affect Based Trust on Behavior

B.2 Mediation of Cognition Based Trust on Behavior

B.3 Mediation of total trust

Behavioral Metric	Model	Model Parameters		
		<i>B</i>	<i>B_{SE}</i>	β
Time to closest approach	Direct Model ($R^2 = .11$)			
	Constant	-0.55	0.07	
	performance	0.02	0.07	.03
	style	0.02	0.07	.03
	interaction	0.16	0.07	.32*
	Mediator Model ($\Delta R^2 = .19\dagger$)			
	Constant	-0.05	0.07	
	performance	0.03	0.07	.06
	style	0.02	0.07	.04
	interaction	0.09	0.07	.19
	total trust	0.16	0.08	.32*
Time to start area	Direct Model ($R^2 = .17^*$)			
	Constant	-0.04	0.09	
	performance	0.00	0.09	.00
	style	0.04	0.09	.06
	interaction	0.27	0.09	.41*
	Mediator Model ($\Delta R^2 = .25^*$)			
	Constant	-0.02	0.09	
	performance	0.02	0.09	.03
	style	0.00	0.09	.00
	interaction	0.18	0.10	.28†
	total trust	0.21	0.10	.31*

Table B.1: Parameter weights for the direct linear models and linear models with affect based trust as the mediator ($\dagger p < .10$, $^* p < .05$, $^{**} p < .01$).

Behavioral Metric	Model	Model Parameters			
		<i>B</i>	<i>B_{SE}</i>	β	
Time to closest approach	Direct Model ($R^2 = .14^{**}$)	Constant	-0.55	0.07	
		performance	0.02	0.07	.03
		style	0.12	0.07	.03
		interaction	0.12	0.07	.32*
	Mediator Model ($\Delta R^2 = .16^*$)	Constant	-0.52	0.07	
		performance	0.08	0.07	.17
		style	0.02	0.07	.04
		interaction	0.09	0.07	.185
		total trust	0.24	0.09	.39**
Time to start area	Direct Model ($R^2 = .11^*$)	Constant	-0.04	0.09	
		performance	0.00	0.09	.00
		style	0.04	0.09	.06
		interaction	0.27	0.09	.41*
	Mediator Model ($\Delta R^2 = .20^{**}$)	Constant	-0.03	0.09	
		performance	0.08	0.09	.12
		style	0.05	0.09	.07
		interaction	0.19	0.09	.28*
		total trust	0.30	0.12	.35*
Start distance	Direct Model ($R^2 = .11$)	Constant	-0.09	0.11	
		performance	0.07	0.11	.08
		style	0.06	0.11	.07
		interaction	0.23	0.11	.30*
	Mediator Model ($\Delta R^2 = .17\dagger$)	Constant	-0.09	0.11	
		performance	0.14	0.11	.18
		style	0.07	0.11	.08
		interaction	0.16	0.12	.20
		total trust	0.28	0.15	.28\dagger
Walking distance	Direct Model ($R^2 = .14\dagger$)	Constant	-0.02	0.14	
		performance	0.04	0.14	.04
		style	-0.04	0.14	-.04
		interaction	0.36	0.14	.36*
	Mediator Model ($\Delta R^2 = .26^{**}$)	Constant	-0.01	0.13	
		performance	0.18	0.14	.18
		style	-0.03	0.14	-.03
		interaction	0.22	0.14	.22
		total trust	0.50	0.19	.40*
Time to closest approach	Direct Model ($R^2 = .05$)	Constant	-0.06	0.07	
		performance	0.02	0.07	.03
		style	0.02	0.07	-.03
		interaction	0.16	0.07	.32*
	Mediator Model ($\Delta R^2 = .16^*$)	Constant	-0.05	0.07	
		performance	0.08	0.07	.17
		style	0.02	0.07	.04
		interaction	0.09	0.07	.19
		total trust	0.24	0.09	.39*
Speed to closest approach	Direct Model ($R^2 = .04$)	Constant	-0.03	0.08	
		performance	0.04	0.08	.06
		style	-0.06	0.08	-.11
		interaction	0.16	0.08	.28\dagger
	Mediator Model ($\Delta R^2 = .09$)	Constant	-0.02	0.08	
		performance	0.09	0.08	.17
		style	-0.06	0.08	-.10

Behavioral Metric	Model	Model Parameters		
		B	B_{SE}	β
Time to closest approach	Direct Model ($R^2 = .11$)			
	Constant	-0.55	0.07	
	performance	0.02	0.07	.03
	style	.012	0.07	.03
	interaction	.912	0.07	.32*
	Mediator Model ($\Delta R^2 = .26 **$)			
	Constant	-1.70	0.55	
	performance	0.08	0.07	.16
	style	0.00	0.06	.00
	interaction	0.06	0.07	.121
total trust	0.34	0.11	.45**	
Time to start area	Direct Model ($R^2 = .17*$)			
	Constant	-0.04	0.09	
	performance	0.00	0.09	.00
	style	0.04	0.09	.06
	interaction	0.27	0.09	.41*
	Mediator Model ($\Delta R^2 = .30 **$)			
	Constant	-2.04	0.71	
	performance	0.08	0.09	.12
	style	0.02	0.08	.03
	interaction	0.15	0.08	.23
total trust	0.42	0.15	.42*	
Start distance	Direct Model ($R^2 = .11$)			
	Constant	-0.09	0.11	
	performance	0.07	0.11	.08
	style	0.06	0.11	.07
	interaction	0.23	0.11	.30*
	Mediator Model ($\Delta R^2 = .18\dagger$)			
	Constant	-1.93	0.91	
	performance	0.14	0.11	.18
	style	0.04	0.11	.05
	interaction	0.12	0.12	.16
total trust	0.38	0.19	.32*	
Walking distance	Direct Model ($R^2 = .14\dagger$)			
	Constant	-0.09	0.11	
	performance	0.04	0.11	.04
	style	-0.03	0.11	-.04
	interaction	0.29	0.11	.36*
	Mediator Model ($\Delta R^2 = .26 **$)			
	Constant	-2.48	0.90	
	performance	0.13	0.11	.16
	style	-0.06	0.11	-.07
	interaction	0.15	0.12	.18
total trust	0.50	0.19	.40*	
Speed to start area	Direct Model ($R^2 = .11$)			
	Constant	-0.38	0.15	
	performance	0.04	0.15	.04
	style	-0.07	0.15	-.06
	interaction	0.32	0.15	.32*
	Mediator Model ($\Delta R^2 = .24*$)			
	Constant	-3.55	1.16	
	performance	0.16	0.14	.16
	style	-0.10	0.14	-.10
	interaction	0.14	0.15	.13
total trust	0.66	0.24	.41**	

Table B.3: Parameter weights for the direct linear models and linear models with total trust as the mediator ($\dagger p < .10$, $* p < .05$, $** p < .01$).

Appendix C

Results for all Participants

Because of the difficulty foreign students had with filling in the questionnaire, it was decided to exclude them from the final analyses as described in Chapter ???. However, some analyses were performed and are included in this Appendix for completeness.

C.1 Differences between Dutch and German participants

Throughout the analyses of the results of the questionnaire, we established that there are significant differences between Dutch and German participants of the experiment. Inclusion of German participants in the results decreased the estimated effect sizes (η^2) sometimes by an order of magnitude. This was caused by higher standard deviations than the Dutch participants. This in turn lowered p-values throughout the analyses. Unfortunately, the German speaking participant group is too small to make reliable statements about the possibility that some of the effects found are substantially different between the Dutch and German groups. However, the fact remains that including them causes less significant results throughout the analysis of the questionnaire. Several explanations of this effect can be given and are stated below.

The first, and most likely explanation, is that the German participants in this study did not master the Dutch language as well as the native Dutch participants. The questionnaire was written in Dutch. When the Germans filled in the questionnaire at the end of the experiment, they were unable to read the items as precise as the Dutch participants. This may account for more spread in their answers.

It is also possible Germans have a different understanding of trust based on cultural differences. Cross-cultural differences have been shown to exist, especially when the cultures are very different from each other (Yuki, Maddux, Brewer, & Takemura, 2005). However, the Dutch and German cultures are closely related, so cultural differences are probably insignificant compared to a linguistic handicap.

For all further sections of this discussion, only the results for Dutch participants are used.

C.2 Analysis of Difference

A 3-way ANOVA which included Nationality as an additional factor was done to see if Nationality had any effect on the total trust score. Only Dutch and German

participants were selected for this test. The 3-way ANOVA with between-subject Performance (bad, good), Style (Shaky, Smooth), and Nationality (Dutch, German) showed a significant Performance \times Style \times Nationality interaction ($F(1, 59) = 2.124$, $p < .05$, $\eta^2 = .07$). Further post-hoc analysis revealed that there was a significant Performance \times Style interaction

Because of these differences between German and Dutch participants found by the 3-way ANOVA described above, the decision was made to exclude foreign participants from subsequent analyses.

C.2.1 Analysis of trust over all participants

A 2-way ANOVA of the total trust score with Performance (bad, good) and Style (shaky, smooth) as between-subject factors showed a marginal significant effect of Performance ($F(1, 64) = 2.941$, $p < .10$, $\eta^2 = .04$). On average, a good performance is more trustworthy ($M = 4.6784$) than a bad performance ($M = 4.9824$). There was also a significant effect of Performance \times Style interaction ($F(1, 64) = 4.391$, $p < .05$, $\eta^2 = .06$). Post-hoc 1-way ANOVAs showed a significant effect of Performance when Style was smooth ($F(1, 64) = 7.481$, $p < .001$, $\eta^2 = .11$). On average, good Performance coupled with a smooth Style scored higher ($M = 5.1259$) than a bad Performance when coupled to a smooth Style ($M = 4.4706$). Also, a marginal significant effect was found of Style in the case Performance was bad ($F(1, 64) = 2.926$, $p < .10$, $\eta^2 = .04$). Interestingly, on average, a shaky Style coupled with a bad Performance scored higher ($M = 4.8863$) than a bad Performance coupled with a smooth Style ($M = 4.4706$).

C.2.2 Differences between cognitive and affectionate trust

In the case of all participants, the ANOVA of cognitive trust showed a marginally significant effect of Performance ($F(1, 64) = 3.767$, $p < 0.10$, $\eta^2 = .06$). On average, robots with a good Performance had a higher cognitive trust score ($M = 5.2614$) than robots with a bad Performance ($M = 4.8529$). No significant effect of Performance \times Style interaction was found ($F(1, 64) = 2.435$, $p > .10$). In contrast, the test for affective trust showed no significant effect of Performance ($F(1, 64) = 0.451$, $p > 0.5$), but a significant Performance \times Style interaction ($F(1, 64) = 4.201$, $p < 0.05$, $\eta^2 = .06$). Post-hoc tests revealed that there was a marginal significant effect of Performance when the robot's Style was smooth ($F(1, 64) = 3.816$, $p < .10$, $\eta^2 = .06$). On average good Performance lead to a higher affectionate trust ($M = 4.7222$) than bad Performance ($M = 4.1667$) when coupled to a smooth Style. Also, a marginal significant difference between Styles was found when coupled to bad Performance ($F(1, 64) = 3.005$, $p < .10$, $\eta^2 = .05$). On average, affectionate trust was higher when a shaky Style was coupled to bad Performance ($M = 4.6667$) than when a smooth Style was coupled to bad Performance ($M = 4.1667$).

List of Tables

1.1	Calibration of Trustworthiness	3
3.1	Descriptives Total Trust	16
3.2	Descriptives Cognition based Trust	17
3.3	Descriptives Affect based Trust	17
3.4	Descriptives Time to gesture	18
3.5	Descriptives Time to closest approach	19
3.6	Descriptives Time to start position	19
3.7	Descriptives Starting distance	19
3.8	Descriptives Walking distance	20
3.9	Descriptives Speed to closest approach	20
3.10	Descriptives Speed to start position	20
3.11	Correlation matrix behavioral variables	21
3.12	Correlations between trust and behavioral variables	22
3.13	Summary of mediation analyses	23
B.1	Mediator models of Affect based trust on Behavior	40
B.2	Mediator models of Cognition based trust on Behavior	41
B.3	Mediator models of Trust on Behavior	42

List of Figures

1.1	Additive and Consistency model of trustworthiness	4
1.2	Model of Trust Processing	5
1.3	TWENDY-one and our robot	7
2.1	Gestures used in the experiment	11
3.1	Graphs of trust metrics	17
3.2	Mediation Model	22