HUMAN-ROBOT TRUST
Is Motion Fluency an Effective Behavioral Style for Regulating Robot Trustworthiness?

Mike Ligthart  Rik van den Brule  Pim Haselager

m.ligthart@student.ru.nl
Radboud University Nijmegen, Montessorilaan 3, 6525 HR Nijmegen

Abstract
Finding good behavioral styles to express robot trustworthiness will optimize the usage of robots. In previous research, motion fluency as behavioral style was studied. Smooth robot motions were compared with trembling robot motions. In a video experiment an effect of motion fluency on trust was found, while in an Immersive Virtual Environment (IVE) experiment, no effect was observed [1]. In this research, we explored the question whether the motion fluency effect is present in a short version of an IVE task and disappears when the task is longer. Results indicate this is not the case. Several explanations for this null-effect are discussed and several recommendations for further human-robot trust studies are provided.

1 Introduction
Social robots are not only the future, they are here right now. Meet TWENDY-ONE (see Figure 1), a humanoid cooperative robot who can help both in the household and in hospitals. It is designed to help ageing people with their daily activities [2]. We use a computer model of TWENDY-ONE in our experiment.

Besides one-on-one human-robot interactions robots are able to improve the performance of human-robot teams by extending human sensory, psychomotor, and cognitive abilities [3]. However, two non-optimal patterns exist, which can be found during experiments as well as real experiences with human-robot teams in the field. The first pattern is when team members overtrust a robot and expect more from the robot than what it actually can perform [4]. This is called misuse. The second pattern is when a team does not trust the robot and chooses not to use it, although the use of the robot would have resulted in a better outcome [5]. This is called disuse. Calibrating trust and aligning the human’s perceived trustworthiness of the robot to reality can avoid misuse and disuse [6]. This means that the calibration of trust influences the success of the interaction between a human and a robot, and plays an important role in the usage of robots in the future [7].

Attempts to regulate trust with robot behavioral styles such as motion fluency, gaze and hesitating behavior are not that promising [1]. After an experiment with video stimuli of a computer-generated robot, the researchers found that only motion fluency (smooth motions versus trembling motions) had an effect on robot trustworthiness. When the same behavioral style was tested in an Immersive Virtual Environment (IVE) experiment, which provides a more realistic experience [8] and the possibility to interact with the robot, no effect of motion fluency was found. It was hypothesized that the duration of the experiment might have influenced the outcome. Because the experiment was longer, participants may have grown accustomed to the behavior of the robot. The researchers hypothesize that motion fluency has an effect on robot trustworthiness in the beginning of
the experiment but that it may have faded away over time. Since trustworthiness was measured with a questionnaire after the complete experiment this would explain why no effect of motion fluency was found [1].

In this paper the possible influence of the duration of the experiment is studied. We try the answer the following question: does the length of the experiment influence the effect of motion fluency on the trustworthiness of a robot?

Two factors were manipulated in the experiment: motion fluency (smooth vs. trembling) and duration (long vs. short). We expect a motion fluency x task duration interaction effect such that there is an effect of motion fluency on trustworthiness only in the short condition of the task. In the long condition, no motion fluency effect on the robot’s trustworthiness is expected. In particular, we expect that in the short duration condition a robot with smooth motions will be valued as more trustworthily than a trembling robot.

2 Method

2.1 Participants

103 participants were recruited from the Radboud University Nijmegen participant pool or signed up after a message on an online social network. As a reward participants received either course credits or a €5 gift card. Three participants were excluded a priori from analysis because they were unable to complete the experiment because of disorientation caused by the IVE, leaving a total of 100 participants (19 men, 81 women, age: median 20.5, range 18-64). An a priori power analysis determined that 100 participants should provide enough power ($\pi = .80$) to detect a medium sized effect ($\eta_p^2 \geq .06$) for both a main effect of motion fluency and an interaction effect of motion fluency x duration [9, 10].

2.2 Immersive Virtual Environment

The experiment was conducted in the Radboud Immersive Virtual Environment Research lab (RIVERlab). The participants wore an nVisor SX60 Head Mounted Display (HMD), which provided participants with a stereoscopic 3D view (frame rate: 60 Hz, resolution: 1280x1024 pixels, horizontal field of view: 44°, vertical field of view: 38°) of the virtual world. On top of the HMD a tracking sensor was placed, which tracked the movements of the participants’ head via an InterSense IS-900 tracking system at a sampling rate of 300Hz. To manipulate a cursor in the virtual world the participants held another tracking sensor in their right hand to track the movements of the hand. The virtual world was generated with WorldViz Vizard 3.0. A standard available virtual room was adapted to suit the needs of the task (see below). The information generated by the tracking system was inserted into Vizard to provide the immersive experience.

2.3 Task

The experiment incorporated two actors; besides the participants, also a virtual robot was present. Both the participants and the robot performed the same task. While the participants performed their own task they had to correct the robot when it made a mistake. The goal of the participants was to maximize the combined score of both tasks.

The task used in experiment was the Van Halen task. It is named after the band Van Halen who presumably wanted a bowl of M&M’s before every show, only the brown M&M’s had to be removed [1]. During the Van Halen task an actor (participant or robot) sat in front of a virtual conveyor belt where balls traversed from right to left. Each ball was assigned a uniform color e.g. blue, red, brown, et cetera. The actor had to pick the brown balls from the conveyor and let the rest pass. For each brown ball taken the actor got a point. More points were collected for every non-brown ball passed. Mistakes, such as removing a non-brown ball, or letting the brown

---

1 Power analysis was performed with G*Power 3.1.7 (http://wwwpsycho.uni-duesseldorf.de/abteilungen/aap/gpower3/)
ball reach the end of the conveyor, were punished by subtracting a point from the player’s total score. The score and mistakes of each player were displayed on a scoreboard.

The first element of the participants’ task was performing the Van Halen task itself. The participants could control a cursor with their right hand, enabling them to remove balls from the conveyor belt. Touching a ball was enough to remove it. The second element of the participants’ task was to correct the robot that also performed the Van Halen task. When the robot made a mistake the participant could prevent the subtraction of points from the robot’s score by pressing a virtual button. When the participant mistakenly pressed the button, i.e. when the robot had not made a mistake, 10 points were subtracted from the participants’ own score and a buzzer sound was played. The used robot avatar was a virtual copy of TWENDY-ONE (see Figure 3). The participants sat in front of their conveyor belt on a rotating chair. The conveyor belts were placed in such a way that they could not be seen in the same view. The participants had to turn to the left on the rotating chair, to check up on the robot (see Figure 2).

2.4 Conditions

Duration and motion fluency were manipulated as between subject factors with each two levels in the experiment. This resulted in four different conditions (see Table 1).

To research if the length of the experiment had an influence on trustworthiness the task duration was manipulated with two levels: long (204s) vs. short (68s). The long duration consisted of three blocks of the original Van Halen task and the short duration consisted of only one block [1]. During each block 48 balls passed the robot and 89 balls passed the participant. Motion fluency (fluently vs. trembling) was also manipulated. The robot accomplished the task respectively in a fluent or trembling manner. In the trembling conditions the arm of the robot vibrated by adding a sine wave to the joints of the robot’s arm and spine joint (7 Hz, amplitude 0.6 degrees) during every grabbing procedure in such a way that it was noticeable for the participants. The robot was still able to pick up a (possible wrong) ball. The performance remained the same in all conditions.

<table>
<thead>
<tr>
<th>Motion Fluency</th>
<th>Duration</th>
<th>Fluently</th>
<th>Trembling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long</td>
<td>Long-Fluently</td>
<td>Long-Trembling</td>
<td></td>
</tr>
<tr>
<td>Short</td>
<td>Short-Fluently</td>
<td>Short-Trembling</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Conditions of the experiment. There were two between subject factors with each two levels. Duration (long vs. short) and motion fluency (fluently vs. trembling).
2.5 Procedure

All participants were escorted from the front desk of the lab facilities to the RIVERlab. After reading the instructions and signing the informed consent form, the participants were asked to take a seat on the rotating chair in the middle of the room. The virtual environment was loaded and the HMD and sensors were placed on the participants’ head and hand respectively. The participants were introduced to the virtual world and the Van Halen task. A practice round was initiated after pressing the virtual green button. Here, the participants could practice their own task: removing brown balls from a conveyor belt. This was followed by the introduction of the robot. The participants were directed to turn left by rotating their chair. Behind the second conveyor belt the robot appeared after the participants pressed the green button. Then, a trial version of the Van Halen task of the robot was initiated. This robot behaved the same for all participants and showed only fluent behavior. The participants were instructed to press the green button whenever the robot made a mistake. After this round, a final practice round was started in which the participants tried both tasks (removing balls from their conveyor and correcting the robot) at the same time. Before the real experiment started, the participants were instructed to look at the robot conveyor belt. After pressing the green button again, the test robot faded out and was replaced by the experiment robot, which had a different color. Participants were told that this second robot was the “real” robot with different characteristics than the “practice” robot. From this point on, the active condition determined the duration of the experiment (short vs. long) and the motion fluency of the new robot (fluently vs. trembling). During the task the experimenter did not instruct the participant any longer but remained present in the lab. When the task was completed, the experimenter took off the HMD and the hand sensor, and the participants filled in a questionnaire containing manipulation checks and robot trustworthiness measures. The experimenter left the room while the participants answered the questions.

2.6 Questionnaire

In this experiment trustworthiness was measured with a questionnaire. Four items were used to capture the trustworthiness of the robot: 1) reliability, 2) trust, a 3) positive and a 4) negative valence item. The negative valence item was reversed-coded for the analysis. The following questions were used:

- **Reliability**: to what extent does the robot appear reliable?
- **Trust**: to what extent would you trust the robot sorting the balls without human supervision?
- **Positive valence**: how positive would you grade the robot?
- **Negative valence**: how negative would you grade the robot?

Responses were recorded on a 7-point Likert scale (1 = not at all, 7 = extremely). These four constructs were averaged in a compound measure of trustworthiness. This compound measure was highly reliable (Cronbach’s \( \alpha = .816 \)). Furthermore, two manipulation checks were included in the questionnaire. One assessed if the participant had seen the robot tremble and the other assessed if the participants perceived a difference in performance (which was not manipulated). The rest of the questions in the questionnaire were for gathering demographics (i.e., gender, age, study program).

3 Results

3.1 Design and Analysis

Unless otherwise specified, all analyses were done with a 2 (duration: long versus short) vs. 2 (motion fluency: smooth versus trembling) between-subject Analysis of Variance (ANOVA). Beforehand, manipulation, experimenter and acquaintance checks were done to test for unwanted effects (see next section).
3.2 Manipulation Checks

3.2.1 Motion Fluency

The manipulation check for the fluency behavior indicated that the participants noticed the motion fluency manipulation. Participants perceived a trembling robot more trembling ($M_{\text{preceived\_trembling}} = 4.00, SD = 1.78$) than a robot with fluent behavior ($M_{\text{preceived\_trembling}} = 1.88, SD = 1.06$): $F(1,96) = 55.08, p < .0001$ and $\eta^2_p = .365$. Although no differences were expected between the long ($M_{\text{preceived\_trembling}} = 2.56, SD = 1.67$) and short ($M_{\text{preceived\_trembling}} = 3.32, SD = 1.88$) condition, a small duration effect was found in the fluency manipulation check: $F(1,96) = 7.08, p = .009$ and $\eta^2_p = .069$. Participants perceived the robot as more trembling in a short condition than in the long condition. No fluency x duration interaction effect was found, $F(1,96) = .490, p = .486$.

3.2.2 Experimenter and Acquaintance Checks

To make sure that the experimenter who guided the experiment had no effect on the trustworthiness measure an experimenter check was done. A three-way ANOVA with trustworthiness as dependent variable and motion fluency, duration and experimenter as between-subject factors confirmed the absence of any influence of the experimenter: all $F$’s < 1.0, $p$’s > .48

Furthermore a one-way ANOVA was conducted to compare trustworthiness when participants respectively knew or did not know one of the experimenters. There was not a significant difference in the scores for knowing the experimenter ($M = 3.15, SD = .99$) and not knowing the experimenter ($M = 3.49, SD = 1.01$) conditions; $F(1,98) = .561, p = .456$.

1.1.1 Performance checks

A performance check was conducted to investigate if participants noticed a difference in performance. The performance measure was the dependent variable where motion fluency and duration were the between subject factors. Since the performance of the robot is kept constant no difference was expected. This was confirmed by a two-way ANOVA: all $F$’s < 1.46, $p$ > .23.

1.2 Trustworthiness

The two-way ANOVA of the trustworthiness measure showed that there was no difference between trembling and fluent behavior in both duration conditions: all $F$’s < 1, $p$’s > .608 (see Figure 4). Meaning that no motion fluency x interaction effect was observed.

Figure 4: the means and 95%-confidence intervals shown in a bar plot. An ANOVA showed no significant difference between any of the conditions.

4 Discussion

The results from the experiment reveal no significant main effect of motion fluency or motion fluency x duration interaction effect on the robot’s trustworthiness. Manipulation checks confirm the participants noticed the trembling of the robot. Experimenter and acquaintance checks confirm the difference in experimenters or participants knowing one of the experimenters did not influence the results. The performance check confirms the assumption that performance is not an interfering factor. With this in mind, there are two possible explanations for the lack of difference between the conditions:
1) There is no effect of motion fluency on trust in both the short and long conditions of the Van Halen task.

2) There is a motion fluency effect on trust, but the experimental design does not allow measuring the effect.

These explanations are discussed in detail in the following sections.

4.1 No effect of Motion Fluency

The first possible explanation is simply that there is no effect at all. If we look at a meta-study of antecedents of robot trust [11] we find that the effect of robot attributes on trust is medium to small compared to other trust related factors such as robot performance or environmental trust factors. Although behavioral styles such as motion fluency are not directly part of the meta-study, they fit in the category of ‘robot attributes’ [11]. Given a power in this study of \( \pi = .80 \) and a sample size of 100 participants, it is reasonable to state that the effect of motion fluency on trust is too small to be detected in this setting or is non-existing.

4.2 Interfering Factors

The lack of difference between all the conditions does not necessarily mean an absence of an effect of motion fluency on trust. In this section we will discuss several possible interfering factors that might hide or counter a motion fluency effect.

4.2.1 Cognitive Load

Why does motion fluency have an effect on trust in the movie experiments and not during the IVE experiments [1]? A big difference is the cognitive load of the task for the participant. In the movie experiments, participants evaluate robot trustworthiness they see in a video. In the IVE experiment, participants make their judgments not only after seeing a robot perform a task, but also after simultaneously performing their own task. In other words during the IVE experiments, participants have less free cognitive resources to actively interpret the possible influences of the trembling of the robot compared to the movie experiments. Although participants do notice the robot’s trembling, this may not become part of their cognitive evaluation concerning the robot’s trustworthiness.

4.2.2 Recency Effect

Measuring trust post-run may mask real-time changes in trust resulting in biases such as the recency effect [12]. This means that when the questionnaire is presented to the participants they recall their most recent experiences with the robot. If this hypothesis is correct and there is an effect of motion fluency in the short condition, we are still too late to measure it, because at this point in the experiment the possible effect of motion fluency is already gone. We have to note that a possibility remains that the short condition was not short enough to measure an effect of motion fluency. However, if the motion fluency effect is that fleeting it would lose its usefulness to regulate robot trustworthiness [12].

4.2.3 Assimilation and Primacy Effect

Participants might not make a big enough distinction between the first robot, meant for introducing the tasks, and the second, experimental robot. Besides the difference in color both robots are equal in appearance. Although it was pointed out to participants that both robots have different behavior, it might not be enough to make a distinction. Participants might project their experience with the first robot onto (the behavior of) the second robot. Since the first robot is constant over all the conditions no difference will be found in the questionnaire. This is called the assimilation effect [13]. A primacy effect might result in the same outcome. Participants use, as opposed to the recency effect, only their first experience for evaluating the robot. Since the first experience was

\[ r = +.03 \]

\[ d = +.47 \]

The correlational effect size (\( r = +.03 \)) is small and the experimental effect size (\( d = +.47 \)) is medium [9, 11]
with the first robot the same consequences as the assimilation effect apply resulting in using the experiences of the introduction robot for evaluating the experiment robot [12].

4.2.4 Effect of Immersive Virtual Environment
Glitches caused by the use of virtual reality, e.g. fluctuations in frame rate that might interfere with the perceiving of the trembling behavior, result in a different experience than was described in the experimental design. In other words, participants might attribute the trembling behavior of the robot (e.g. trembling) as an artifact of the IVE setup instead of attributing it directly to the robot itself.

5 Recommendations
This study contributes to the field of social robots by indicating that there is no interaction effect between motion fluency and the duration of the task and by formulating recommendations that improves the quality of future research. Taking into account both recent trust measuring techniques and widely accepted concepts, such as assimilation effects and cognitive load, improvements in experimental setups can be made.

5.1 Cognitive Load
The cognitive load of a task might influence the way participants assess robot behavior. Clearly more research is needed. As a first, general rule of thumb, we suggest that the higher the cognitive load, the less aspects of a robot’s behavioral style will affect human trust in the robot. This means that for researching potentially weak trust factors a task with a low cognitive load is required. On the other hand most realistic scenarios using these factors are in a condition where there is a relatively high load involved.

5.2 Real-time Trust Measurement
Measuring trust real-time can prevent biases such as the primacy-recency effect to have an influence on the outcome. Early drops in reliability of traditional post-run survey approaches are found [12]. By prompting short trust questions during a task, or by obtaining information on the development of trust otherwise, the evolution of trust during tasks is acquired. This can be represented as a trust curve that shows the amount of trust over time. A trust value is then calculated by taking the area under the trust curve (AUTC) [12].

5.3 Increasing Contrast
Increasing the contrast between two objects can prevent an assimilation effect [13]. In our case this would mean using different models for each of the robots. In other words the difference in appearance should be made more distinct, yet care has to be taken that this would not lead to a confounding factor.

5.4 Real Robots
Participants may attribute (parts of) the behavior of the robot to Immersive Virtual Environment. Glitches during the experiments might have interfered with the participants’ evaluation of the robot. The use of real robots, although presenting different problems, evades this complication, and in general an increase of trust observed [14]. If a study uses an IVE setup, e.g. due to financial or practical reasons, care should always be given to create a contrast between behavior caused by the virtual environment and the stimuli itself.

6 Conclusion
The investigation of motion fluency on a robot’s trustworthiness in this paper shows that an effect of motion fluency is observed neither after a short nor long interaction with a robot. This implies that our hypothesis, that a motion fluency effect can be expected after a short task an thus leads to a motion fluency x duration interaction effect, can be rejected. On a more positive note: this research project increases our understanding of the number of factors that influences development of trust between humans and robots. Cognitive load, primacy/recency effects, assimilation/contrast effects and the use of an Immersive Virtual Environment may all play a role in the human understanding and interpretation of robot behavior. Although trust is a vital element in the future employment of social robots, currently it is clear that we are only at the beginning of our understanding of the complexity and dynamic nature of human robot interaction.
References


