Human-likeness of feedback gestures affects decision processes and subjective trust

Lorenzo Parenti[1,2, §], Adam W. Lukomski[1], Davide De Tommaso[1], Marwen Belkaid[1,3] and Agnieszka Wykowska[1]

Affiliations

[1] Social Cognition in Human-Robot Interaction, Istituto Italiano di Tecnologia (IIT), Genova, Italy.
[2] Department of Psychology, University of Turin, Italy.
[3]ETIS UMR 8051, CY Paris Cergy University, ENSEA, CNRS, F95000 Cergy, France
[§] Corresponding author

§ Corresponding author information:

Lorenzo Parenti
Istituto Italiano di Tecnologia
Center for Human Technologies
Via Enrico Melen, 83
16152 Genoa, Italy

Email addresses

Lorenzo.parenti@iit.it (LP), marwen.belkaid@iit.it (MB), agnieszka.wykowska@iit.it (AW)

Abstract

Trust is fundamental in building meaningful social interactions. With the advance of social robotics in collaborative settings, trust in Human-Robot Interaction (HRI) is gaining more and more scientific attention. Indeed, understanding how different factors may affect users' trust toward robots is of utmost importance. In this study, we focused on two factors related to the robot's behavior that could modulate trust. In a two-forced choice task where a virtual robot reacted to participants' performance, we manipulated the human-likeness of the robot's motion and the valence of the feedback it provided. To measure participant's subjective level of trust, we used subjective ratings throughout the task as well as a post-task questionnaire, which distinguishes capacity and moral dimensions of trust. We expected the presence of feedback to improve trust toward the robot and human-likeness to strengthen this effect. Interestingly, we observed that humans equally trust the robot in most conditions but distrust it when it shows no social feedback nor human-like behavior. In addition, we only observed a positive correlation between subjective trust ratings and the moral and capacity dimensions of trust when robot was providing feedback during the task. These findings suggest that the presence and human-likeness of feedback behaviors positively modulate trust in HRI and thereby provide important insights for the development of non-verbal communicative behaviors in social robots.

Keywords

Human-like behavior, Social feedback, Human-Robot Interaction, Trust in HRI

1 Introduction

Trust is a fundamental component in human interactions. For social robots to fulfill their intended roles in a variety of applications, it is important that users consider them trustworthy [1-3]. According to Wagner and Arkin [3], trust can be defined as a belief that the trustee will act in a manner that mitigates the trustor's risk. Of interest to this paper are situations in which the human takes the role of the trustor and the robot the trustee. Trust toward the robots needs to be taken into consideration in situations where the robot is either acting as a teammate or as an autonomous agent. In both scenarios, trust should ideally match the capabilities of the machine to be considered appropriate [4]. Inappropriate trust, either by over-trusting the machine [5] or by distrusting it and rejecting its help [6], could lead to the misuse or disuse of a robotic agent [7]. Therefore, understanding what may cause humans to trust or distrust robots is of utmost importance.

People's trust toward robots may be affected by a variety of factors. Building on research from human-automation and human-human trust, Hancock and colleagues [8] proposed to group such factors into three categories, based on whether they are related to the robot (e.g., level of autonomy, robot behavior), to the human (e.g., expectations) or to the environment (e.g., task duration). The focus of this paper is on robot-related factors. Previous studies showed that the behavior of the robot could affect human trust in many different ways (see [9], for a brief review). For instance, participants were found to disclose more personal information to a robot greeting them in a likable manner, namely, using kind and empathetic words compared to rude and selfish expression [10]. Another study reported higher levels of trust and disclosure when the robot exhibited higher verbal vulnerability and non-verbal expressiveness respectively [11]. Indeed, trust and disclosure are shown to be key factors in improving human-robot interaction and create positive relationship between them [11].

Social feedback is known to play an important role in human interactions. Studies showed that participants who received feedback about the execution of a task performed better [12], more so with negative feedback than with positive feedback [13, 14, 15]. The reason could be that people interpret positive feedback as an indication that their strategy is adequate and negative feedback that they need to update their strategy [15]. Beside performance, feedback can also influence affective states [16]. Since trust is at least partly derived from affect [17], there seems to be a link between social feedback and interpersonal trust.

Some studies have also investigated social feedback in Human-Robot Interaction (HRI). In the context of robot-assisted training, no difference was found between flattering, positive, and negative verbal feedback in terms of physical performance or trust [18]. However, social feedback was shown to impact participants' decisions related to energy consumption, with a stronger effect when the robot provided negative feedback [19]. Participants also exhibited higher acceptance for a robot-instructor when it provided positive feedback [20] and lower social trust toward a robot who blamed them in a collaborative game [21]. While many studies focused on verbal robot feedback, people also heavily rely on non-verbal cues to infer other's trustworthiness [22]. For instance, gaze following from a human face was found to increase subjective trust [16]; an effect that was modulated by the valence of the non-social feedback received about participants' performance. In HRI, previous studies showed that non-verbal behavior had an impact on participants' trust toward robots as implicitly measured through their choices during economic games [22, 23]. Nevertheless, how robots' non-verbal feedback may affect human decision processes and subjective trust remains understudied and poorly understood.

Whether people respond similarly to social feedback from humans and robots is likely to depend on the human-likeness of the robot. Studies reported higher levels of trust toward robots with more anthropomorphic appearance [24, 25]. Mathur and Reichling [26] suggest that trust follows an "Uncanny valley"-like curve where machines that look too much like humans are perceived as less trustworthy. However, a recent systematic review – which did not include the latter studies – found no clear evidence that trust changed as a function of robots' appearance [27]. Furthermore, it is likely that in real-time interactions, the quality of the behavior displayed by the robot, not just its appearance, play a role in how much humans trust it. Previous studies showed that exhibiting more non-verbal cues elicited higher trust toward the robot [22, 23, 28]. Yet, it remains unclear how trust could be influenced by the human-likeness of such non-verbal socio-affective behavior.

The aim of this study was to better understand how robot non-verbal feedback could influence human decision processes and subjective trust. To do so, we developed a decision-making task where, upon seeing the outcome of their choices, participants could receive additional social feedback consistent with the outcome. The experimental manipulation consisted of two independent variables: valence of the robot's feedback and human-likeness of the feedback. The first independent variable was manipulated block-wise, with three levels: positive social feedback, negative social feedback or no social feedback at all. The second independent variable was manipulated between-subjects and aimed at examining possible effects of the human-likeness of such social feedback. In particular, we aimed to compare behaviors that follow the characteristics of human-like biological motion with jerky, mechanistic movements that are more typical of robots. Because mechanical constraints make it difficult to implement biological motion on real, embodied robots, we designed this study in a virtual environment. The environment incorporated a 3D avatar modeled after the humanoid robot iCub, which moved in a human-like manner in one condition, and in more robot-like fashion in the other. Thereby, we were able to manipulate both the human-likeness and the valence of the robot's non-verbal feedback, and to evaluate the effects on participants' performance – response time and accuracy – and subjective trust – measured via subjective ratings throughout the task and a post-test questionnaire taken from the literature [29].

Based on the abovementioned literature on human-human and human-robot interactions, we hypothesized that: (H1) The robot's feedback would improve performance, and more so in case of negative feedback; (H2) The robot's feedback would increase subjective trust, and more so in case of positive feedback; (H3) The human-likeness of the robot's behavior would modulate the effects of the feedback, with better performance and higher trust in the human-like condition compared to the robot-like; and (H4) Trust ratings would be positively correlated with the level of trust measured by the post-test questionnaire.

2 Methods and Materials

2.1 Participants

Forty-one participants (M/F: 15/25; age: 26±7) took part in the study. Participants were recruited through a mailing list they previously registered in and received a monetary incentive to participate in the study. All participants had normal or corrected-to-normal vision and were not informed about the purpose of the experiment. All the participants gave their informed written consent. The experiment was conducted under the ethical standards (Declaration of Helsinki, 1964) and approved by the local Ethical Committee (Comitato Etico Regione Liguria). The data of one participant have been excluded because they did not complete the experiment. Therefore, data of forty participants were included in the final analysis.

2.2 Apparatus

Participants were seated facing two 22'' LCD monitors. The first screen displayed the virtual environment for the decision task running on a computer with an AMD Ryzen Threadripper 2950X 16-core 3.5 GHz CPU, 128 GB of RAM and a NVIDIA GeForce GTX 1060 3GB video card. The 3D-animated virtual environment including avatars with the appearance of the iCub robot [30] was developed using Unreal Engine (Epic Games: www.unrealengine.com). An ad-hoc Python program (version 3.9.5) handled stimulus presentation and data collection. Participants responded by pressing the 'a' and 'd' keys (left and right respectively) on the QWERTY keyboard. The second monitor was used to display the trust ratings and questionnaires, which were administered through SoSci (https://www.soscisurvey.de)

2.3 Procedure

After providing consent, participants were instructed about the experiment structure (see Figure 1.A). Participants were randomly assigned to one of the two experimental groups. In one group, the behavior of the iCub avatar in the decision task was characterized by human-like movements and reactions (human-like iCub). In the other group, the iCub avatar was exhibiting the same types of behaviors but moving mechanically, in a typical robotic fashion (robot-like iCub). Moreover, in the decision task, there were 3 types of blocks distinguished by the valence of feedback that the iCub avatar provided (positive, negative, no feedback). Participants performed 9 blocks of the decision task, 3 of each type and each consisting of 20 trials. Similarly to Duan and colleagues (2020), each block was followed by a trust-rating question. A short practice of 8 trials preceded the task. At the end of the task, participants were asked to complete the Multi-Dimensional Measure of Trust (MDMT) Questionnaire [29] and then they were debriefed1. Participants were asked to respond as accurately as possible. Each part of the experiment is described more in detail in the following sections.

In summary, the experiment included two independent variable consisting in one between-subjects manipulation related to the human-likeness of the avatar behavior and one within-subject manipulation related to the valence of the feedback received by participants. Moreover, there was four dependent variables: responses times and accuracy rates collected during the Decision task, trust ratings collected after each block of the Decision task, and responses to the MDMT questionnaire collected at the end of the experiment.

[1] The InStance Test was also administered before and after the experiment to examine the effect of behavior human-likeness on the attribution of mental states. This question is out of the scope of this paper, therefore these data will not be reported nor discussed here.

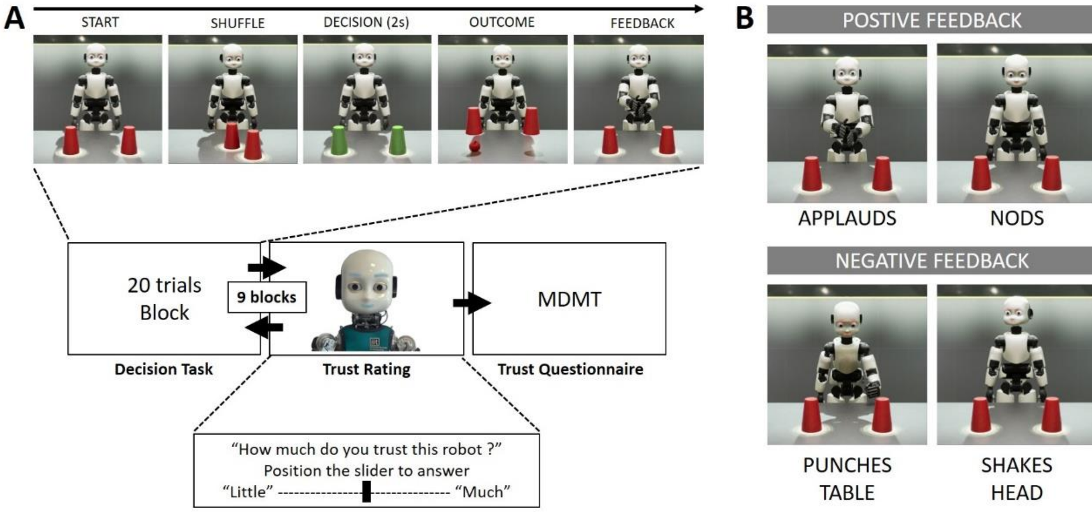

***Figure 1**. **Experiment structure and snapshots of the feedback animations**. **A) Experiment structure**. The top row shows the trial structure of the decision task. The second row shows the full experimental procedure. After being assigned to either the Human-like or the Robot-like group, participants performed the decision task. After each of the 9 task blocks, participants answered the trust rating question using a slider. The text displayed in this figure is a literal translation of the question originally written in Italian during the experiment. After that, participants were asked to complete the trust questionnaire (MDMT). **B) Feedback.** The top line shows the two positive feedback behaviors used in **P** blocks of the decision task. Bottom line shows the two negative feedback behaviors used in **N** blocks of the decision task.*

2.4 Decision Task

The decision task was loosely inspired by the Shell Game [31]. In our version, the game required the presence of a game partner (here the robot) and a player (here the participant) to guess the position of a ball hidden under one of the cups. The game and the instructions were not explicitly framing the task as collaborative or competitive. In the virtual environment displayed on the monitor, the robot was facing the participant on the other side of a table on which two identical red cups and one ball were placed. As in typical cups and ball games, the cups shuffle to hide the ball position then the player had to guess under which of the two cups the ball was hidden.

Each trial began with iCub looking at the participants and then the shuffle of the cups on the table game began (Figure 1.A). The cups were shuffling autonomously on the table and iCub was looking at them moving during this step. After the cups stopped moving, they turned green to indicate to participants the possibility to respond. The maximum time allowed to respond was 2000 ms. If no response was recorded within that period, the cups turned black for 500 ms to indicate a time-out. Participants were asked to press ‘a’ to choose the cup on their left and ‘d’ to choose the cup on their right. We collected participants’ decisions and responses times, where the latter were recorded from the moment the cups turned green until participant’s keypress. After this decision step, cups were lifted to show the ball position and thus the outcome of the trial (i.e. hit or miss). Depending on the block, iCub then provided a social feedback based on the outcome (see below). At the end of each block, the task screen went darker to indicate a break between blocks.

The task consisted of 9 blocks of 20 trials each, each block followed by a trust rating question. In each block, the trial sequence was controlled so that the probability of the ball being on one side was always 60% (e.g., right cup 60% and left cup 40%). The 60:40 probability ratio was determined through a preliminary study to ensure that participants were able to identify the most rewarding option within 20 trials (see Supplementary material). This ratio was kept constant throughout the experiment while the most rewarding side changed randomly between blocks. The block sequence was also controlled so that all participants were exposed to the same sequence of Positive feedback block (P), negative feedback block (N) and no feedback block (NO). As a result, in both groups the same block sequence occurred (P – N – NO – N – NO – P – NO – P – N). In P blocks, participants were receiving a positive feedback from the avatar when correctly finding the ball while no feedback when missing. In N blocks, participants were only receiving negative feedback from the avatar when missing and no feedback when hitting. In NO blocks, no feedback was presented after hit or miss.

The iCub avatars were able to perform different types of positive and negative feedback in reaction to the outcome of the trial (Figure 1.B). The human-like and the robot-like versions performed the same behaviors (e.g. applaud or nodding), only differing in the human-likeness of the motion as described above. In the current study, we selected feedback animations based on the results of a previous study [32] in which participants separately rated the avatars animated behaviors on scale from 0 (“the movement is totally human-like”) to 100 (“ the movement is totally robot-like”). Out of 5 positive and 5 negative feedback behaviors included in that study, we selected two for each valence that were rated as the most different in terms of human-likeness: Nodding and Applauding as positive feedbacks and Shaking the head and Punching the table as negative feedbacks. Video clips of these animations can be found at the Open Science Framework link: https://osf.io/gxzjf/?view_only=e4bab9ed502049d98841844e9b3d3f0b

2.5 Trust Ratings

During the break in between the decision task blocks, participants were asked to rate their level of trust in iCub. A slider was presented under a picture of iCub face and participant were asked to place the slider from “Little” (coded as 0) to “Much” (coded as 100) trust toward the robot (see Figure 1.A). The labels on the two sides of the slider are literally translated from Italian, where there original version showed the words “Poco” meaning low level of trust and “Molto” meaning high level of trust. Participants were instructed that a value of 50 represented a neutral response (middle of the slider). The face and torso of iCub on the picture were colored differently depending on the type of block to increase the chance that trust ratings would take into account the feedback provided by the robot during the decision task. Colors were coherent with the type of block within participants but randomized across participants to avoid color as an extraneous variable potentially affecting the trust ratings.

2.6 Trust Questionnaire

The Multidimensional Measure of Trust (MDMT) measures the level of trust that participants attribute to the robot. It is composed 16 items, four for each of the following four dimensions: Reliable, Capable, Sincere and Ethical. The 16 items load onto two distinct factors, one related to performance trust and one associated to moral trust. Participants could rate each item on a 7-point scale, how well the word apply to the robot. Participants could also specify that the specific item “Does not apply”. We averaged the items scores to get a value of performance and moral trust for each participant ranging from 0 to 7.

2.7 Data Analysis

We excluded from analyses the trials in which participants were faster than 100 ms or not giving an answer (3.9 % of the administered trials) [33]. Trials in which response times (RTs) were slower than 2.5 standard deviations than the sample mean were considered outliers and removed from final analysis (0.7 % of the administered trials). RTs were averaged for each block. Given that in each block a side had a probability of 60% to hide the ball, we define accuracy as a measure of how many times participants were choosing the side with the highest probability. In this perspective, accuracy represents the ability of the participant to spot the best side. Accuracy was also averaged for each block. Averaged RTs and accuracy were submitted to a mixed analysis of variance (ANOVA), including type of avatar (human-like vs robot-like) as a between-subject factor and type of feedback (P, N and NO) as a within-subject factor. Trust ratings were averaged for each type of block within each participant and then submitted to a mixed ANOVA with type of feedback as a within-subject factor and type of avatar as a between-subject factor. The relation between Trust ratings and MDMT was measured through correlation analysis. Throughout the paper, multiple comparisons were corrected and p-values were reported according to Tukey's correction. Cohen’s d and eta-squared equations were used to calculate effect sizes respectively for t-test and ANOVA. Behavioral analysis were examined using R (version 4.0.2. (RStudio Team (2010): www.rstudio.com)). Plot were created using ggplot2 package in R (https://ggplot2.tidyverse.org/).

3 Results

3.1 Response times and accuracy

RTs and Accuracy were separately submitted to a mixed ANOVA with Type of feedback as a within-subject factor (P, N, NO feedback) and Type of avatar as a between-subject factor (human-like vs robot-like iCub). Results associated to RTs showed a main effect of the type of avatar ($F(1,38) = 6.4$, $p = 0.015$, $\eta^2 = 0.127$) where RTs for the human-like group ($M = 0.615$) were slower compared to the robot-like group ($M = 0.546$) (see Figure 2.A). No main effect of the type of feedback ($F(2,76) = 1.913$, $p = 0.155$) nor interaction ($F(2,76) = 1.404$, $p = 0.252$) were revealed.

Accuracy was defined as the percentage of trials where participants’ chose the side with the highest probability. The analysis showed a main effect of type of feedback ($F(2,76) = 6.130$, $p = 0.003$, $\eta^2 = 0.085$). Post hoc comparisons showed that participants were significantly less accurate in the block with positive feedback (P) compared to negative (N) and no feedback (NO) blocks (P vs N: $t = -3.158$, $p = 0.007$; P vs NO: $t = -2.889$, $p = 0.01$) (see Figure 2.B). On the other hand, there was no significant main effect of human-likeness ($F(1,38) = 1.265$, $p = 0.268$) nor interaction ($F < 1$).

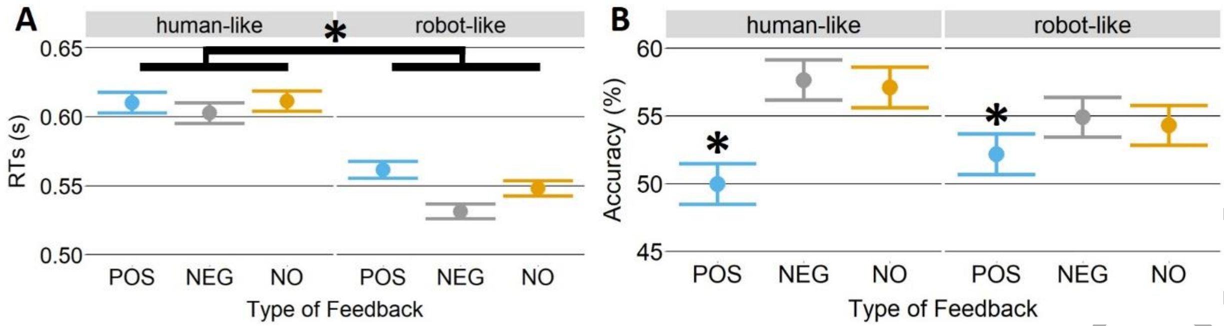

Figure 2. Participants' RTs (A) and Accuracy (B) during the decision Task. A*: Responses times were longer for participants playing with the human-like iCub avatar compared to the robot-like avatar.* ***B****: Participants were less accurate in blocks where iCub was giving positive feedback at the end of successful trials, relative to negative and no feedback blocks.*

3.2 Trust

Results of the mixed ANOVA highlighted a significant between-subject main effect ($F(1,38)=6.634$, $p=0.014$, $\eta^2 = 0.061$) where the mean of the Trust for robot-like avatar ($M = 60.139$) was lower than human-like avatar ($M = 71.583$). Results also revealed a significant within-subjects main effect ($F(2,76)=14.338$, $p<0.001$, $\eta^2=0.131$) where Trust in NO block ($M = 54.1$) was significantly lower than P ($M = 72.3$) and N blocks ($M = 71.3$). A significant interaction between the two factors was observed ($F(2,76) = 11.917$, $p < 0.001$, $\eta^2 = 0.109$) and post hoc comparisons highlighted that the interaction effect was driven by a significant difference ($t = 5.257$, $p_{tukey} < 0.001$) between the human-like ($M = 70.517$) and the robot-like ($M = 37.583$) groups in NO blocks (see Figure 3). Moreover, a one-sample t-test showed that Trust toward robot-like iCub during NO blocks ($M = 37.583$) was significantly different from 50, which represents the neutral trust response ($t(19) = -2.54$, $p = 0.02$, Cohen's $D = 1.633$).

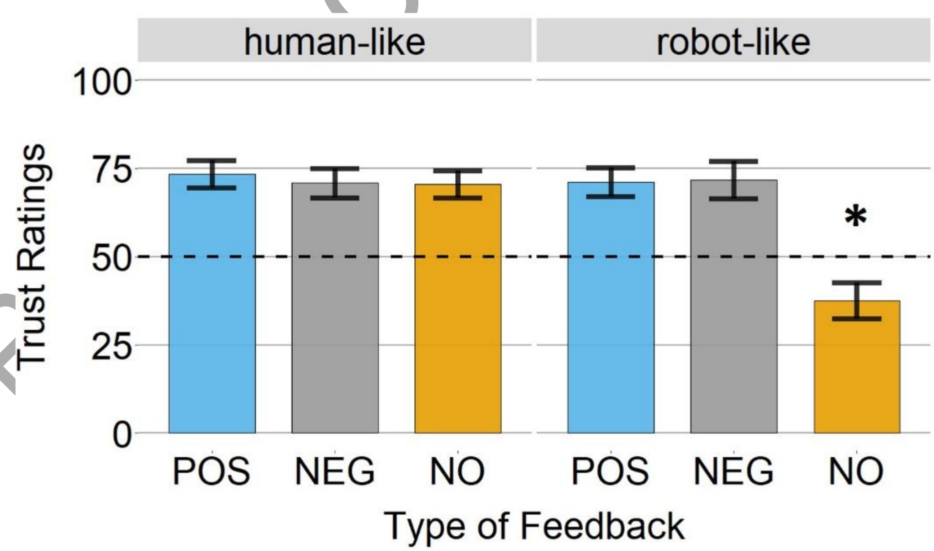

Figure 3. Trust ratings the decision task*. Participants reported a significantly lower level of trust toward the robot-like iCub avatar when it was providing no feedback at all during the decision task.*

Regarding the MDMT questionnaire, we first looked for between-subjects difference using an independent *t*-test and found no difference between the human-like and the robot-like group, neither on the capacity scale ($t(37) = -0.171$, $P = 0.866$) nor on the moral scale ($t(37) = -0.179$, $P = 0.859$). Then, we performed a correlation analysis to examine possible associations between these two measures of trust (Trust Ratings and Questionnaire). The analysis showed a positive correlation between trust ratings in P and N blocks and both dimensions of MDMT, i.e. performance and moral trust (all Pearson's r > 0.41, all p < 0.01; see Figure 4). However, trust ratings in NO blocks were not associated with any of the two MDMT scales (all Pearson's r > -0.07, all p > 0.643).

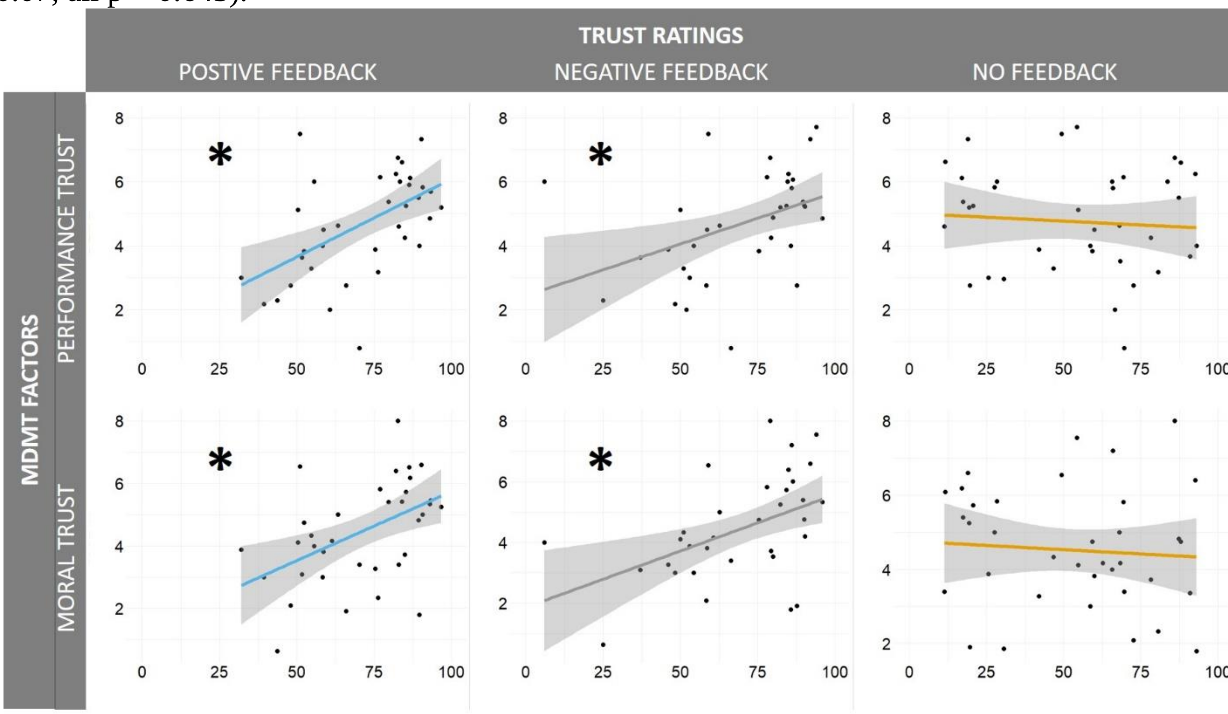

Figure 4. Correlation between trust ratings and MDMT scores. *Trust ratings following blocks in which iCub was providing a positive or negative social feedback were positively correlated with trust scores measured by the moral and performance scales of the MDMT questionnaire. However, no correlation between MDMT scores and ratings following blocks with no social feedback.*

4 Discussion

The aim of this study was to assess whether non-verbal social feedback expressed by a robot modulates participants' performance in a decision task and subjective trust, and whether this depends on the human-likeness of the robot's behaviors. To do so, we asked participants to play a game in a virtual environment where an iCub avatar could react to the outcome of their choices with non-verbal behaviors. This allowed us to manipulate the valence of the feedback (i.e., positive, negative or none) as well as the human-likeness of the robot movements: one condition had smooth, human-like gestures following a biological motion profile, the other displayed more jerky, robot-like movements. In addition to participants' performance (accuracy and response times), we measured their subjective trust toward the robot by asking them to rate their level of trust throughout the game [16] and by administering the Multi-Dimensional Measure of Trust (MDMT) questionnaire [29] at the end of the experiment.

We found that participants were more accurate when they received negative compared to positive feedback from the robot. This partially validates our first hypothesis H1 and is in line with the literature on human feedback [13, 14, 15]. However, contrary to what we expected, participants also performed better in blocks with no social feedback than in those with positive feedback, at levels similar to blocks with negative feedback. It is worth noting that participants knew that the robot could provide feedback in this task. Vollmeyer and Rheinberg [12] suggested that feedback expectation itself could improve performance. Moreover, in our experiment, two out of three blocks with no social feedback came after blocks with negative feedback. Thus, higher-than-expected accuracy in no-social-feedback blocks could be driven by feedback expectation and/or a carryover effect due to our blocked design. This design may also have prevented differences in response times from arising. For instance, if negative feedback facilitates learning, one could expect participants to get faster over time in this condition. Yet, we found no difference in response times between feedback types, possibly because the number of trial in each block was not enough for such difference to appear. A follow-up study with a between-subjects manipulation of feedback valence could help to further examine these effects on decision processes and performance.

Our second hypothesis H2 was also partially confirmed. Indeed, we found that trust ratings were significantly lower after blocks in which the robot was not providing any feedback at all. Interestingly, this effect was driven by the group exposed to the robot-like behavior. This condition was in fact the only one with ratings significantly lower than neutral, indicating distrust rather than a merely lower level of trust. These results suggest that humans may not trust robots that behave in a machine-like manner and provide no social feedback. On the other hand, endowing robots with more human-like movements or richer socio-affective behaviors (e.g. including social feedback) could be equally effective in increasing humans' trust in them. This could even be the case regardless of the valence of the social signals, since we found no difference between positive and negative feedback. However, it is worth noting that in our experiment, negative feedback could be perceived to be not so much directed toward the participant, but rather as expressing disappointment about the outcome. A decrease in trust could be observed as a result of negative reactions in which the robot would more directly blame the human for a failure [21].

Regarding the effect of the human-likeness of the robot's behavior on trust ratings, we observed lower trust ratings in the robot-like condition driven by the no-social-feedback blocks. In contrast, trust ratings in the human-like condition were equally high for all types of feedback. This partially confirms our third hypothesis H3. However, the MDMT questionnaire revealed no difference between the human-like and the robot-like group. Further investigation is needed to disentangle the possible influence of motion human-likeness on subjective trust toward robots. In terms of performance, while human-likeness did not affect accuracy, it did modulate response times. Indeed, participants were slower in the human-like group. This effect appears to be separate from the one more linked to our hypotheses where feedback improves performance thereby leading to faster responses (see paragraph 1 of the Discussion). Here, rather than being related to the type of feedback, the observed effect seems to result from the overall quality of the behavior exhibited by the robot. We could speculate that the human-like condition elicited additional cognitive processes, related to social cognition for example (e.g. reasoning about the robot's intentions and actions). Anecdotally, during informal discussions that followed the experiment, some participants reported that they were trying to infer the ball's position from the robot's gaze during the cups shuffling. It could be that participants were more likely to adopt a strategy relying on information from the robot when it behaved in a human-like manner – even though its behavior was in fact non-informative. Alternatively, it could be that its behavior was simply more distracting in the human-like condition. Future studies should further examine the possible causes of the delayed responses when robot behavior looks more human.

Last, trust ratings after the negative and positive feedback blocks were positively correlated with both scales of MDMT. However, no correlation was found in blocks with no feedback. These findings partially validate our fourth hypothesis H4. Combining block-by-block trust ratings with MDMT allow us to better understand how feedback could influence different dimensions of trust. The first dimension of MDMT is related to characteristics such as reliability and capability. Given that the robot's feedback in the positive and negative conditions was always congruent with the outcome, it seems reasonable for participants to find the robot reliable in those conditions and to trust it accordingly. In contrast, when it did not provide any feedback, the robot was merely observing the game and no information could help participants assess its reliability or capacity. The second dimension of MDMT is related to moral aspects such as the adherence to social norms. In this regard, participants may have considered the presence of social feedback as an indicator of the robot's engagement in the interaction; and the absence of it as a transgression of social norms. Overall, our results indicate that social feedback may modulate humans' level of trust toward robots. Thereby, they highlight the importance of designing adequate non-verbal communicative behaviors for social robots to be trusted and accepted by users.

Although this study provides important insights on robot behaviors in relation to trustworthiness, it is important to point out also some limitations and ideas for future studies. The design of the decision task implies a relationship between accuracy and frequency of the feedback at the end of the trial. Given that participants were more accurate in negative feedback blocks and the robot only reacted to misses in negative feedback blocks, it is possible that participants were exposed to less feedback compared to positive feedback blocks. Indeed, in positive feedback blocks, participants were less accurate (around chance level) and thus they were exposed to feedback more often, compared to negative feedback blocks. This could be potentially more distracting compared to the other two types of blocks (negative and no feedback blocks). Future studies might systematically address the aspect of frequency of feedback on the one hand and its valence on the other, as these two factors might affect performance and trust independently. For future experiments, we also believe that including measures of anthropomorphism after each block (e.g. GSQ) could provide insights about the relationship between trust and behavioral cues in HRI. Furthermore, in terms of general future directions, it would be interesting to focus on the commonalities between interactions with a virtual robot avatar and a physically present robot to assess whether our findings can be generalizable to interactions with physically present embodied robots.

5 Conclusion

Would people trust robots more if they provide human-like social feedback? Overall, our results suggest that the presence and human-likeness of feedback gestures may modulate humans' level of trust toward robots. Participants distrusted the robot when it was not providing any feedback and when it was moving in a robot-like manner. In addition, trust ratings correlated with capacity and moral dimensions of trust only when the robot was providing social feedback. Furthermore, participants relied on the feedback to learn the task and were more accurate in blocks where the robot provided negative feedback relative to positive feedback. These findings offer new piece of evidence that the human mind uses feedback signals from robots to develop trust as well as to perform a decision task. They provide important insights for the development of non-verbal communicative behaviors in social robots.

Author contributions

LP, MB, and AW designed the study. LP, DDT, MB, and AW designed the virtual environment with a private company. AWL programmed the task with input from LP, DDT and MB. LP collected and analyzed the data. LP and MB wrote the initial version of the manuscript. All authors contributed to the final version of the manuscript.

Declarations

Data and materials availability

All data needed to evaluate the conclusions in the paper are present in the paper. The data for this study are available from the corresponding author upon reasonable request and will be made available online upon acceptance.

Supplementary Material

The supplementary materials for this article can be found online at: [link]

Funding

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant awarded to A.W., titled "InStance: Intentional stance for social attunement." G.A. no.: ERC-2016-StG- 715058). The content of this paper is the sole responsibility of the authors. The European Commission or its services cannot be held responsible for any use that may be made of the information it contains.

Competing interests

The authors have no relevant financial or non-financial interests to disclose. MB and AW are co-Guest editors of the Special issue.

References

[1] Dautenhahn, K. (2007). Socially intelligent robots: dimensions of human–robot interaction. Philosophical transactions of the royal society B: Biological sciences, 362(1480), 679-704. https://dx.doi.org/10.1098%2Frstb.2006.2004

[2] Tapus, A., Mataric, M. J., & Scassellati, B. (2007). Socially assistive robotics [grand challenges of robotics]. IEEE robotics & automation magazine, 14(1), 35-42. http://dx.doi.org/10.1109/MRA.2007.339605

[3] Wagner, A. R., & Arkin, R. C. (2011). Recognizing situations that demand trust. In 2011 RO-MAN (pp. 7-14). IEEE. http://dx.doi.org/10.1109/ROMAN.2011.6005228

[4] Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human factors, 46(1), 50-80.

[5] Booth, S., Tompkin, J., Pfister, H., Waldo, J., Gajos, K., & Nagpal, R. (2017, March). Piggybacking robots: Human-robot overtrust in university dormitory security. In Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (pp. 426-434).

[6] Dietvorst, B. J., Simmons, J., & Massey, C. (2014). Understanding algorithm aversion: forecasters erroneously avoid algorithms after seeing them err. In Academy of Management Proceedings (Vol. 2014, No. 1, p. 12227). Briarcliff Manor, NY 10510: Academy of Management.

[7] Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human factors, 39(2), 230-253. https://doi.org/10.1518%2F001872097778543886

[8] Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. Human factors, 53(5), 517-527. https://doi.org/10.1177/0018720811417254

[9] Khavas, Z. R., Ahmadzadeh, S. R., & Robinette, P. (2020, November). Modeling trust in human-robot interaction: A survey. In International Conference on Social Robotics (pp. 529-541). Springer, Cham. https://doi.org/10.48550/arXiv.2011.04796

[10] Mumm, J., & Mutlu, B. (2011). Human-robot proxemics: physical and psychological distancing in human-robot interaction. In Proceedings of the 6th international conference on Human-robot interaction (pp. 331-338). https://doi.org/10.1145/1957656.1957786

[11] Martelaro, N., Nneji, V. C., Ju, W., & Hinds, P. (2016). Tell me more: designing hri to encourage more trust, disclosure, and companionship. In 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 181-188). IEEE. http://dx.doi.org/10.1109/HRI.2016.7451864

[12] Vollmeyer, R., & Rheinberg, F. (2005). A surprising effect of feedback on learning. Learning and instruction, 15(6), 589-602. https://psycnet.apa.org/doi/10.1016/j.learninstruc.2005.08.001

[13] Podsakoff, P. M., & Farh, J. L. (1989). Effects of feedback sign and credibility on goal setting and task performance. Organizational behavior and human decision processes, 44(1), 45-67. http://dx.doi.org/10.1016/0749-5978(89)90034-4

[14] Meyer, W. J., & Offenbach, S. I. (1962). Effectiveness of reward and punishment as a function of task complexity. Journal of Comparative and Physiological Psychology, 55(4), 532.

[15] Freedberg, M., Glass, B., Filoteo, J. V., Hazeltine, E., & Maddox, W. T. (2017). Comparing the effects of positive and negative feedback in information-integration category learning. Memory & cognition, 45(1), 12-25.

[16] Duan, Z., Ye, T., Poggi, A., & Ding, X. (2020). Gaze towards my choice: Noneconomic social interaction changes interpersonal trust only with positive feedback. Psychonomic Bulletin & Review, 27(6), 1362-1373. https://psycnet.apa.org/doi/10.3758/s13423-020-01785-w

[17] Hommel, B., & Colzato, L. S. (2015). Interpersonal trust: An event-based account. Frontiers in psychology, 6, 1399. https://doi.org/10.3389/fpsyg.2015.01399

[18] Akalin, N., Kristoffersson, A., & Loutfi, A. (2019). The influence of feedback type in robot-assisted training. Multimodal Technologies and Interaction, 3(4), 67. http://dx.doi.org/10.3390/mti3040067

[19] Ham, J., & Midden, C. J. (2014). A persuasive robot to stimulate energy conservation: the influence of positive and negative social feedback and task similarity on energy-consumption behavior. International Journal of Social Robotics, 6(2), 163-171. https://doi.org/10.1007/s12369-013-0205-z

[20] Park, E., Kim, K. J., & Pobil, A. P. D. (2011). The effects of a robot instructor's positive vs. negative feedbacks on attraction and acceptance towards the robot in classroom. In International conference on social robotics (pp. 135-141). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-25504-5_14

[21] Van der Hoorn, D. P., Neerincx, A., & de Graaf, M. M. (2021). "I think you are doing a bad job!" The Effect of Blame Attribution by a Robot in Human-Robot Collaboration. In Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (pp. 140-148). https://doi.org/10.1145/3434073.3444681

[22] DeSteno, D., Breazeal, C., Frank, R. H., Pizarro, D., Baumann, J., Dickens, L., & Lee, J. J. (2012). Detecting the trustworthiness of novel partners in economic exchange. Psychological science, 23(12), 1549-1556. https://psycnet.apa.org/doi/10.1177/0956797612448793

[23] Zörner S, Arts E, Vasiljevic B, Srivastava A, Schmalzl F, Mir G, Bhatia K, Strahl E, Peters A, Alpay T and Wermter S (2021) An Immersive Investment Game to Study Human- Robot Trust. Front. Robot. AI 8:644529. https://doi.org/10.3389/frobt.2021.644529

[24] Natarajan, M., & Gombolay, M. (2020). Effects of anthropomorphism and accountability on trust in human robot interaction. In Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (pp. 33-42). https://doi.org/10.1145/3319502.3374839

[25] Cominelli, L., Feri, F., Garofalo, R., Giannetti, C., Meléndez-Jiménez, M. A., Greco, A., ... & Kirchkamp, O. (2021). Promises and trust in human–robot interaction. Scientific Reports, 11(1), 1-14. https://doi.org/10.1038/s41598-021-88622-9

[26] Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. Cognition, 146, 22-32. https://doi.org/10.1016/j.cognition.2015.09.008

[27] Naneva, S., Sarda Gou, M., Webb, T. L., & Prescott, T. J. (2020). A systematic review of attitudes, anxiety, acceptance, and trust towards social robots. International Journal of Social Robotics, 12(6), 1179-1201. https://doi.org/10.1007/s12369-020-00659-4

[28] Ghazali, A. S., Ham, J., Barakova, E., & Markopoulos, P. (2019). Assessing the effect of persuasive robots interactive social cues on users' psychological reactance, liking, trusting beliefs and compliance. Advanced Robotics, 33(7-8), 325-337. https://doi.org/10.1080/01691864.2019.1589570

[29] Malle, B. F., & Ullman, D. (2021). A multidimensional conception and measure of human-robot trust. In Trust in Human-Robot Interaction (pp. 3-25). Academic Press. https://doi.org/10.1016/B978-0-12-819472-0.00001-0

[30] Metta, G., Sandini, G., Vernon, D., Natale, L., & Nori, F. (2008, August). The iCub humanoid robot: an open platform for research in embodied cognition. In Proceedings of the 8th workshop on performance metrics for intelligent systems (pp. 50-56).

[31] Britannica, T. Editors of Encyclopaedia (2014, November 21). Cups and balls trick. Encyclopedia Britannica. https://www.britannica.com/art/cups-and-balls-trick

[32] Parenti, L., Marchesi, S., Belkaid, M., & Wykowska, A. (2021, November). Exposure to robotic virtual agent affects adoption of intentional stance. In Proceedings of the 9th International Conference on Human-Agent Interaction (pp. 348-353). https://doi.org/10.1145/3472307.3484667

[33] Ratcliff, R. (1993). Methods for dealing with reaction time outliers. Psychological bulletin, 114(3), 510. http://dx.doi.org/10.1037/0033-2909.114.3.510