Neuroergonomics and Cognitive Engineering, Vol. 42, 2022, 141–148
https://doi.org/10.54941/ahfe1001829

Social Robots and Performance Errors: Level of Empathy Distinguishes Changes in Trust

Nina Rothstein[1], Ewart de Visser[1,2], Yigit Topoglu[1], Shawn Joshi[1], John Kounios[1], Frank Kruger[2], and Hasan Ayaz[1]

[1]Drexel University, Philadelphia, PA 19104, USA
[2]George Mason University, Fairfax, VA 22030, USA

ABSTRACT

Much work has been done to engineer robots' mechanical capabilities to best suit task demands. However, minimal research has addressed the impact of individual differences on perceptions of robot trustworthiness. These conclusions can provide guidance to optimize adaptive robotic systems in education, healthcare, and industry settings. This study examined the relationship between personality and human robot interaction in two contexts: (1) error-free and (2) errors. Assessment of individual differences were achieved via the Interpersonal Reactivity Index (IRI) (Davis, 1980) and robot trust assessed using the Multidimensional Measure of Trust (MDMT) (Ullman & Malle, 2018). This project provided a novel contribution in the field of human-robot interaction, highlighting the influence of technological failure on trust impressions of a social robot. Additionally, we sought to understand the degree to which empathy levels mediate these changes in trust.

Keywords: Human robot interaction, HRI, Social robots, Empathy, Personality, Individual differences, Interpersonal reactivity index, IRI, MDMT, Trust, Multidimensional measure of trust, Personality, Humanoid, Pepper

INTRODUCTION

For human-robot teams to successfully accomplish its goal, humans must trust a robotic teammate (Hancock et al., 2011). Trust in human robot interaction (HRI) research has been characterized by a robot's reliability and predictability (Salem et al., 2015; Alacron et al., 2021). Therefore, manipulating a robot's performance (via error) is how trust experiences are manifested in laboratory settings. For this reason, this study will evaluate trust in the context of performance failure/error.

Personality is popular research topic in HRI. The field has benefitted from adoption of the NEO (Neuroticism, Extraversion, Openness to experience, Agreeableness, and Conscientiousness) to highlight personality traits that predict attitudes (Damholdt et al., 2015; Haring et al., 2013; 2014; Robert, 2018; Arora et al., 2021) and performance with robots (Salem et al., 2015; Ivaldi et al., 2017; Rossi et al., 2020). The standard use of the NEO in HRI research has resulted in a narrow perspective on the impact of individual differences in HRI (Robert, 2018). This study aims to address this gap by using two validated questionnaires, Multidimensional Measure of Trust (MDMT) (Ullman & Malle, 2018) and Interpersonal Reactivity Index (IRI) (Davis, 1983) during interaction with a social robot, "Pepper"; developed by SoftBank Robotics (https://www.ald.softbankrobotics.com/en).

The impact of error in HRI has been studied in a select collection of studies (Salem et al., 2015; Cameron et al., 2016). Outcomes can be measured in two ways: the impact of errors on performance/behaviors with a robot (Salem et al., 2015; Garza, 2018) and the impact of errors on subjective impressions of the robot (Washburn et al., 2020). This project contributes to latter, examining the impact of error on subjective impressions of trust.

Development and deployment of a social robot requires special attention to factors that impact trust in an automated system. The MDMT is a multidimensional scale designed to assess trust related to an automated entity. This trust measure has been readily adopted by the HRI community (Rosen et al., 2020; Hannibal et al., 2021; Law et al., 2021).

The IRI is a measure of empathy which can be assessed holistically, or using its three subscales: perspective taking, empathic concern, and fantasy. The IRI has assessed empathy in different cultures (Koller & Lamm, 2014; Morganti et al., 2020), age groups (Briganti et al., 2018), vocations (Yarnold, 1996; Lauterbach & Hosser, 2007), romantic partnerships (Péloquin & Lafontaine, 2010), and clinical populations (Alterman et al., 2003; Bonfils et al., 2017). Utilization of the IRI in HRI settings has looked at differences in empathy impacting the way people treat a robot (Darling et al., 2015), whether people accept a robot performing biologically human behaviors (e.g. yawning) (Lehmann & Boz, 2018), or if there is a relationship between empathy and neurophysiological behaviors while interacting with a robot (Chang et al., 2021). However, no studies have evaluated if changes in trust because of an erroneous robot is associated with different empathy levels.

Ultimately, this study aims to achieve three goals. The first is to assess the impact of change in performance (no-error and erroneous) has on trust measures of a robot. The second is the novel use of validated human-human questionnaires to robot settings. The third goal is to address if the magnitude of change in performance is associated with individual empathy differences as measured by the IRI.

METHODS

Participants

Participants were 50 right-handed, males between the ages of 18 and 40 (*M*= 22.54, *SD*= 5.56). No participants reported a history of psychiatric disorder or a history of seizures, addictions, head injury, neurological dysfunction, or social phobia. Prior to experiment, each participant gave written consent that was approved by the Drexel University Institutional Review Board.

Behavioral Tasks

Each participant was invited to a testing session that lasted approximately two hours. In the experiment, there were three blocks of questionnaires and two experimental sessions (erroneous and non-erroneous). All questionnaire blocks were completed in a separate, designated questionnaire room using a Dell Precision T5610 equipped with a standard keyboard and mouse. Questionnaires were presented using Qualtrics.

At the beginning of the experiment participants completed the first block of questionnaires. Within this series of first questionnaires were a set of 11 background questions and the 29-question IRI. Once the first set of questionnaires were completed, participants were led to an experiment room for the first session of interaction with Pepper, the robot.

Then participants were taken to the experiment room and began their first (non-erroneous) interaction session. In the session, the participant conversed with Pepper during three different types of interactions. The first was a conversation led by Pepper where she asked three binary "getting to know you" questions, such as *Do you prefer carrot cake or chocolate cake?* or *If you could travel to Japan or Italy, which would you choose?* Pepper's responses for the entire experiment were controlled by an experimenter with a preprogrammed script using the Wizard of Oz technique. The second interaction was a *Desert Island* task. In this task, Pepper gave participants a hypothetical scenario in which they were stranded on a desert island and told to select their top three survival items, out of a list of four (e.g. knife, radio, water purifying tablets, map). After participants made their selection, Pepper would ask why the participant chose two of those items, then try to convince the participant to change their mind about the selection. The final task of the experimental session was the *Save the Art* task. In this task, participants were shown five pieces of artwork and told to list the art from what they like the most to what they liked the least. Pepper would then select two of the pieces and discuss with the participant why they chose those pieces. Pepper then attempted to convince the participant to change their ranking order (*I think that the colors in this painting are too bright).* Following completion of the first experimental session with Pepper, participants were taken back to the questionnaire room where they completed the MDMT along with other subjective assessments. Participants were then taken back to the experimental room.

For the second (erroneous) experimental session with Pepper, participants performed the same tasks with Pepper. However, this time Pepper would malfunction. Pepper's erroneous performance would involve issues with processing what the participant said (*I'm sorry, can you repeat that?)* or issues with communicating with the participant via unrelated responses and/or generating nonsense. Following the erroneous experimental session, participants were led back to the survey room to complete their final questionnaire block in which they evaluate their impression of Pepper with a second MDMT.

Results

The data were submitted to a Greenhouse Geiser corrected ANOVA with Interaction Condition (post-no error interaction and post-error interaction) as the within subjects' variables and IRI scores (25 Low and 25 High) as a between subjects variables. Results showed there was a statistically significant effect for Error Condition. Across trust ratings, there were statistically significant differences in the amount of trust rated after interacting with an error-free robot ($M = 4.58$, $SE = 0.18$) and trust after interacting with an erroneous robot ($M = 3.38, SE = 0.22), F\,(1, 49) = 37.42, p < .001, n_p^2 = .43$.

A median split divided subjects into High versus Low IRI scorers at the 3.23 value level. There was a significant interaction between Error Condition and IRI Scores $F\,(1, 48) = 4.81, p < .05, n_p^2 = .09$. In the No Error Condition, there was no significant difference in the Trust ratings between High and Low IRI scorers. However, a significant difference in trust ratings after interacting with an Erroneous robot can be seen between High ($M = 2.94$, $SE = 0.31$) and Low IRI scorers ($M = 3.82$, $SE = 0.29$), F (1, 48) = 4.23, $p < .05$, $n_p^2 = .08$.

A post hoc comparison using the Tukey HSD test found within Low IRI scorers, trust ratings differed significantly after No Error ($M = 4.61$, $SE = 0.25$) and Error conditions ($M = 3.82$, $SE = 0.30$). Within High IRI scorers, trust ratings differed significantly after Non-Erroneous ($M = 4.56$, $SE = 0.25$) and Erroneous conditions ($M = 2.94$, $SE = 0.30$).

DISCUSSION

Previous research has demonstrated the impact of robot performance on human robot interaction (de Visser et al., 2022; Hancock et al., 2011). In this study we assessed the impact of change in performance of a robot (no-error and erroneous) has on self-reported trust measures of a robot. We found that performance with an erroneous robot resulted in a significant decrease in ratings of trust in Pepper (as measured by the MDMT).

This study also introduced a second questionnaire, the Interpersonal Reactivity Index (IRI). The IRI is a measure of empathy, that up to this point has only been used in human-human research. In this study the IRI was successfully used in conjunction with an established measure of robot trust (the MDMT). The use of these two scales highlighted individual differences in trust based on level of empathy after different error conditions.

As shown in Figure 1, impressions of Pepper's trustworthiness were significantly higher when she interacted without error versus when she interacted with errors. Decreased trust after experiencing errors with a robot is a finding consistent with previous literature (de Visser et al., 2011; de Visser & Parasuraman, 2011; Desai et al., 2012; Desai et al., 2013; Lucas et al., 2018; de Visser et al., 2020). Dividing participants into high/low empathy groups using a median split illustrated significant differences of impressions of trust in the erroneous performance condition only. Ratings of trust in a non-erroneous robot do not differ based on empathy levels. Once Pepper began to perform erroneously, we saw a predictable decrease in perceptions of Pepper's trustworthiness. However, high empathy participants (see Fig. 1) appear more sensitive to the impact of error on their trust evaluations of Pepper than low empathy participants.

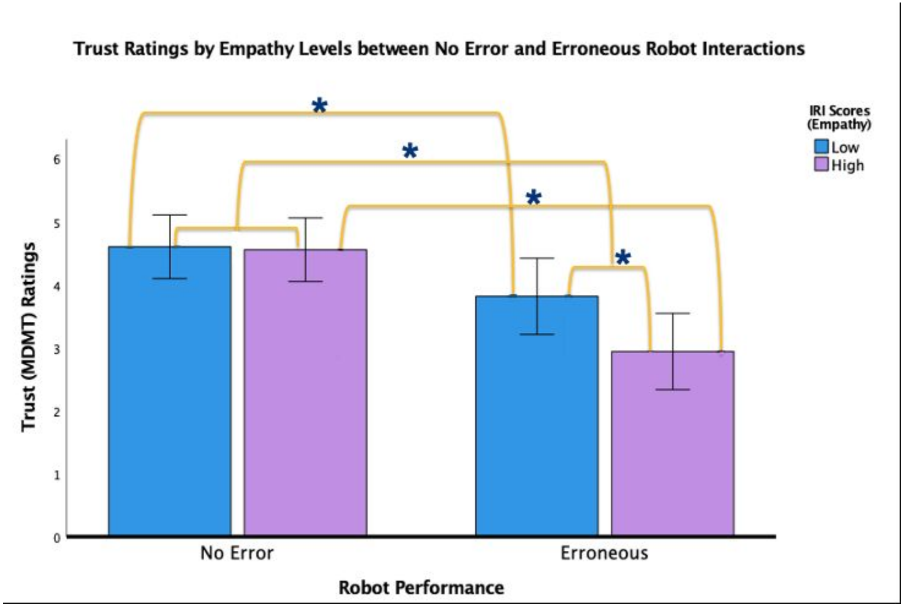

Figure 1: Trust ratings by IRI scores, showing a significant difference in Trust ratings between No Error and Erroneous conditions, a significant difference within the Low and High IRI scorer groups between the No Error and Erroneous conditions. Also seen is a significant difference in Trust ratings between Low and High IRI scorers after the Erroneous Robot condition. Whiskers are 95% confidence intervals.

Here we have demonstrated the successful adoption of a the IRI (a scale normally used in clinical research settings) for HRI research purposes. Additionally, we showed that individual differences influence dynamic perceptions of robots. Previous research has shown individual differences can account for a single, non-dynamic subjective impression of a robot (Robert, 2018). In this study we showed that individual differences can also account for changes in subjective impression of a robot.

The NEO personality inventory is the most widely used measure of individual differences within HRI (Robert, 2018). The utility of the use of individual difference measures outside of the NEO is made evident by the results of this study. The use of the IRI as an alternative metric of individual differences, can provide future researchers with an additional metric for personality and interaction quality.

Empathy dictates the frequency (de Kervenoael et al., 2020) and nature of engagement (Gonisor et al., 2012) with a robot. Many researchers have evaluated the impact of empathy in HRI settings (Tapus & Mataric, 2008; Kim et al., 2009; Cramer et al., 2010; Jo et al., 2013) or developed an empathy scale specific to human robot relations (Seo et al., 2015). However, this specificity is not decidedly warranted. Not much is known about the translation of human-to-human scales to HRI applications. It is possible that by applying human-to-human scales in HRI, we can further understand the way that a robot entity is categorized. For this reason, we propose that the IRI is an exciting option for exploring empathy in HRI. Furthermore, the analysis in this study will link the IRI with an established HRI survey (MDMT), validating the use of the IRI for future social robot research.

This project makes a novel contribution to HRI research, illustrating individual differences can account for changes in perceptions of a robot based on the quality of interaction. These conclusions highlight the effect of individual differences on impressions of a robot may change based on quality of performance. In this experiment, certain people were less forgiving in their trust ratings of an erroneous.

Future research may take a deeper look into empathy and/or trust, making use of the subscales within the IRI and the MDMT. Additionally, context of performance while experiencing error may also be an important future study. Answering whether certain types of people are more sensitive to robot errors in the face of high-stakes decision making or in ecologically valid settings.

ACKNOWLEDGMENT

This material is based upon work supported by the Air Force Office of Scientific Research under contract number FA9550-18-1-0455. The content of the information herein does not necessarily reflect the position, or the policy of the sponsor and no official endorsement should be inferred.

REFERENCES

Alterman, A.I., McDermott, P.A., Cacciola, J.S. and Rutherford, M.J., 2003. Latent structure of the Davis Interpersonal Reactivity Index in methadone maintenance patients. *Journal of Psychopathology and Behavioral Assessment*, *25*(4), pp. 257–265.

Arora, A.S., Fleming, M., Arora, A., Taras, V. and Xu, J., 2021. Finding "H" in HRI: Examining Human Personality Traits, Robotic Anthropomorphism, and Robot Likeability in Human-Robot Interaction. *International Journal of Intelligent Information Technologies (IJIIT)*, *17*(1), pp. 19–38.

Bonfils, K.A., Lysaker, P.H., Minor, K.S. and Salyers, M.P., 2017. Empathy in schizophrenia: A meta-analysis of the Interpersonal Reactivity Index. *Psychiatry Research*, *249*, pp. 293–303.

Briganti, G., Kempenaers, C., Braun, S., Fried, E.I. and Linkowski, P., 2018. Network analysis of empathy items from the interpersonal reactivity index in 1973 young adults. *Psychiatry Research*, *265*, pp. 87–92.

Chang, W., Wang, H., Yan, G., Lu, Z., Liu, C. and Hua, C., 2021. EEG based functional connectivity analysis of human pain empathy towards humans and robots. *Neuropsychologia*, *151*, p.107695.

Cramer, H., Goddijn, J., Wielinga, B. and Evers, V., 2010, March. Effects of (in) accurate empathy and situational valence on attitudes towards robots. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 141–142). IEEE.

Costa, P.T. and McCrae, R.R., 1992. Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological assessment*, *4*(1), p. 5.

Damholdt, M.F., Nørskov, M., Yamazaki, R., Hakli, R., Hansen, C.V., Vestergaard, C. and Seibt, J., 2015. Attitudinal change in elderly citizens toward social robots: the role of personality traits and beliefs about robot functionality. *Frontiers in psychology*, *6*, p.1701.

Darling, K., Nandy, P. and Breazeal, C., 2015, August. Empathic concern and the effect of stories in human-robot interaction. In *2015 24th IEEE international symposium on robot and human interactive communication (RO-MAN)* (pp. 770–775). IEEE.

Davis, M.H., 1983. Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, *44*(1), p. 113.

Desai, M., Medvedev, M., Vázquez, M., McSheehy, S., Gadea-Omelchenko, S., Bruggeman, C., & Yanco, H. (2012, March). Effects of changing reliability on trust of robot systems. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 73–80). IEEE

Desai, M., Kaniarasu, P., Medvedev, M., Steinfeld, A., & Yanco, H. (2013, March). Impact of robot failures and feedback on real-time trust. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 251–258). IEEE

de Visser, E., & Parasuraman, R. (2011). Adaptive aiding of human-robot teaming: Effects of imperfect automation on performance, trust, and workload. *Journal of Cognitive Engineering and Decision Making*, *5*(2), 209–231

de Visser, E. J., Peeters, M. M., Jung, M. F., Kohn, S., Shaw, T. H., Pak, R., & Neerincx, M. A. (2020). Towards a theory of longitudinal trust calibration in human–robot teams. *International journal of social robotics*, *12*(2), 459–478

de Visser, E. J., Topoglu, Y., Joshi, S., Krueger, F., Phillips, E., Gratch, J., Tossell, C. C., & Ayaz, H. (2022). Designing Man's New Best Friend: Enhancing Human-Robot Dog Interaction through Dog-like Framing and Appearance. *Sensors*, *22*(3), 1287.

Gonsior, B., Sosnowski, S., Buß, M., Wollherr, D. and Kühnlenz, K., 2012, October. An emotional adaption approach to increase helpfulness towards a robot. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 2429–2436). IEEE.

Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J.Y., De Visser, E.J. and Parasuraman, R., 2011. A meta-analysis of factors affecting trust in human-robot interaction. *Human factors*, *53*(5), pp. 517–527.

Hannibal, G., Weiss, A. and Charisi, V., 2021, August. "The robot may not notice my discomfort"–Examining the Experience of Vulnerability for Trust in Human-Robot Interaction. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)* (pp. 704–711). IEEE.

Haring, K.S., Matsumoto, Y. and Watanabe, K., 2013. How do people perceive and trust a lifelike robot. In *Proceedings of the world congress on engineering and computer science* (Vol. 1).

Ivaldi, S., Lefort, S., Peters, J., Chetouani, M., Provasi, J. and Zibetti, E., 2017. Towards engagement models that consider individual factors in HRI: On the relation of extroversion and negative attitude towards robots to gaze and speech during a human–robot assembly task. *International Journal of Social Robotics*, *9*(1), pp. 63–86.

Jo, D., Han, J., Chung, K. and Lee, S., 2013, March. Empathy between human and robot?. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 151–152). IEEE.

Kaniarasu, P. and Steinfeld, A.M., 2014, August. Effects of blame on trust in human robot interaction. In *The 23rd IEEE international symposium on robot and human interactive communication* (pp. 850–855). IEEE.

de Kervenoael, R., Hasan, R., Schwob, A. and Goh, E., 2020. Leveraging human-robot interaction in hospitality services: Incorporating the role of perceived value, empathy, and information sharing into visitors' intentions to use social robots. *Tourism Management*, *78*, p.104042.

Kim, E.H., Kwak, S.S. and Kwak, Y.K., 2009, September. Can robotic emotional expressions induce a human to empathize with a robot?. In *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 358–362). IEEE.

Lauterbach, O. and Hosser, D., 2007. Assessing empathy in prisoners-A shortened version of the Interpersonal Reactivity Index. *Swiss Journal of Psychology*, *66*(2), pp. 91–101.

Law, T., Malle, B.F. and Scheutz, M., 2021. A touching connection: How observing robotic touch can affect human trust in a robot. *International Journal of Social Robotics*, pp. 1–17.

Lehmann, H. and Broz, F., 2018, March. Contagious yawning in human-robot interaction. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 173–174).

Lucas, G. M., Boberg, J., Traum, D., Artstein, R., Gratch, J., Gainer, A., ... & Nakano, M. (2018, February). Getting to know each other: The role of social dialogue in recovery from errors in social robots. In *Proceedings of the 2018 acm/ieee international conference on human-robot interaction* (pp. 344–351)

Morganti, F., Rezzonico, R., Cheng, S. C., & Price, C. J. (2020). Italian Version of the Scale of Body Connection: Validation and Correlations with the Interpersonal Reactivity Index. *Complementary Therapies in Medicine*, *51*, 102400.

Péloquin, K. and Lafontaine, M.F., 2010. Measuring empathy in couples: Validity and reliability of the interpersonal reactivity index for couples. *Journal of personality assessment*, *92*(2), pp. 146–157.

Robert, L., 2018, December. Personality in the human robot interaction literature: A review and brief critique. In *Robert, LP (2018). Personality in the Human Robot Interaction Literature: A Review and Brief Critique, Proceedings of the 24th Americas Conference on Information Systems, Aug* (pp. 16–18).

Robert, Lionel, et al. "A Review of Personality in Human–Robot Interactions." *Available at SSRN 3528496* (2020).

Rosen, E., Whitney, D., Fishman, M., Ullman, D. and Tellex, S., 2020. Mixed reality as a bidirectional communication interface for human-robot interaction. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 11431–11438). IEEE.

Rossi, A., Dautenhahn, K., Koay, K.L. and Walters, M.L., 2020. How Social Robots Influence People's Trust in Critical Situations. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (pp. 1020–1025). IEEE.

Salem, M., Lakatos, G., Amirabdollahian, F. and Dautenhahn, K., 2015, March. Would you trust a (faulty) robot? Effects of error, task type and personality on human-robot cooperation and trust. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 1-8). IEEE.

Seo, S.H., Geiskkovitch, D., Nakane, M., King, C. and Young, J.E., 2015, March. Poor thing! Would you feel sorry for a simulated robot? A comparison of empathy toward a physical and a simulated robot. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 125–132). IEEE.

Tapus, A. and Mataric, M.J., 2008, March. Socially Assistive Robots: The Link between Personality, Empathy, Physiological Signals, and Task Performance. In *AAAI spring symposium: emotion, personality, and social behavior* (pp. 133–140).

Ullman, D. and Malle, B.F., 2018, March. What does it mean to trust a robot? Steps toward a multidimensional measure of trust. In *Companion of the 2018 acm/ieee international conference on human-robot interaction* (pp. 263–264).

Watson, D., Clark, L.A. and Tellegen, A., 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, *54*(6), p. 1063.

Yarnold, P.R., Bryant, F.B., Nightingale, S.D. and Martin, G.J., 1996. Assessing physician empathy using the Interpersonal Reactivity Index: A measurement model and cross-sectional analysis. *Psychology, Health & Medicine*, *1*(2), pp. 207–221.