



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Transparency in HRI: Trust and Decision Making in the Face of Robot Errors

Citation for published version:

Nesset, B, Robb, DA, Lopes, J & Hastie, H 2021, Transparency in HRI: Trust and Decision Making in the Face of Robot Errors. in *HRI '21 Companion: Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. ACM Association for Computing Machinery, United States, pp. 313–317, 16th ACM/IEEE International Conference on Human-Robot Interaction, Virtual Conference, 9/03/21. <https://doi.org/10.1145/3434074.3447183>

Digital Object Identifier (DOI):

[10.1145/3434074.3447183](https://doi.org/10.1145/3434074.3447183)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

HRI '21 Companion: Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.





Transparency in HRI: Trust and Decision Making in the Face of Robot Errors

Birthe Nasset, David A. Robb, José Lopes, Helen Hastie
 Heriot-Watt University
 Edinburgh, U.K.
 bn25,d.a.robb,jd.lopes,h.hastie@hw.ac.uk

ABSTRACT

Robots are rapidly gaining acceptance in recent times, where the general public, industry and researchers are starting to understand the utility of robots, for example for delivery to homes or in hospitals. However, it is key to understand how to instil the appropriate amount of trust in the user. One aspect of a trustworthy system is its ability to explain actions and be transparent, especially in the face of potentially serious errors. Here, we study the various aspects of transparency of interaction and its effect in a scenario where a robot is performing triage when a suspected Covid-19 patient arrives at a hospital. Our findings consolidate prior work showing a main effect of robot errors on trust, but also showing that this is dependent on the level of transparency. Furthermore, our findings indicate that high interaction transparency leads to participants making better informed decisions on their health based on their interaction. Such findings on transparency could inform interaction design and thus lead to greater adoption of robots in key areas, such as health and well-being.

CCS CONCEPTS

• **Human-centered computing** → **Natural language interfaces**; *Graphical user interfaces*; • **Computer systems organization** → *Robotic autonomy*.

KEYWORDS

Social robotics, transparency, trust, robot errors, hospital triage.

ACM Reference Format:

Birthe Nasset, David A. Robb, José Lopes, Helen Hastie. 2021. Transparency in HRI: Trust and Decision Making in the Face of Robot Errors. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21 Companion)*, March 8–11, 2021, Boulder, CO, USA. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3434074.3447183>

1 INTRODUCTION

As AI and robotic systems become more complex and autonomous, making decisions on their own, users become less aware of their intent and internal processes. Furthermore, if users are hesitant to adopt them and do not trust them appropriately, they will be unable to make use of their full potential. This is particularly so



This work is licensed under a Creative Commons Attribution-ShareAlike International 4.0 License.

HRI '21 Companion, March 8–11, 2021, Boulder, CO, USA.

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8290-8/21/03. <https://doi.org/10.1145/3434074.3447183>

in the healthcare domain, where understanding rationale behind decisions and actions is key. If appropriate levels of transparency can be designed into health diagnostic systems, then such systems can be made more effective, will be more readily adopted and thus result in freeing up of crucial health resources [16].

In this paper, we use a robot performing COVID-19 triage to investigate how transparency and the degree to which a robot makes errors affects user trust and decision making. Specifically, we deliberately probe whether there is an accompanying effect on potential patients' concrete decisions about whether or not to seek additional human advice, beyond what the robot is telling them. We use this to explore if the *appropriate* amount of trust has been instilled in the user because if the robot makes an error (as sometimes they and their human counterparts do), then they would likely ask for a second opinion. Although, we do acknowledge that people's propensity to trust does vary from person to person [6, 8].

As we had restricted access to human subjects for this study, we used video vignettes, which is a method used in previous studies on HRI interaction and shown to be effective [3, 19, 26]. The aim of the study is to show how different levels of transparency influence the users' trust towards a system and their decision making. Through manipulating the level of system transparency and the level of system error, we set out to discover how to optimise transparency for human-robot interaction. The contributions are thus as follows:

- Empirical evidence that high transparency can help users of robotic diagnostics to calibrate their trust level and react appropriately, based on information provided through human-robot interaction.
- Empirical evidence that system transparency and system errors interact together to affect user trust in the system.
- This empirical evidence in this hospital triage context adds to prior work on the effect of transparency and errors on trust in other domains.

2 BACKGROUND

In this section, we discuss briefly the area of transparency, human-robot trust and Theory of Mind.

Transparency and Theory of Mind. Transparency is seen as a desirable attribute for autonomy for both functional and ethical reasons [2]. The EPSRC UK research council, in its principles for robotics states that: "*Robots are manufactured artefacts. They should not be designed in a deceptive way to exploit vulnerable users; instead their machine nature should be transparent.*" [4]. Bhaskara et al. define transparency as aiming "*to provide operators an awareness of an autonomous agent's behaviour, reliability, and intention*" [1]. Lyons' models of transparency suggest that for the interaction between a system and a user to be effective, the user needs to

understand information about the system (robot-to-human transparency) and the system needs to understand information about the user (robot-of-human transparency) [1, 12]. Furthermore, robot-of-human transparency requires the system to have some awareness of human participation in the task and to understand the user's beliefs, desires and intentions. This has also been formalised as Theory of Mind [15] and has previously been applied to HRI [22], and has indeed been inextricably linked to trust [20, 21]. Theory of Mind has also been stipulated as one of the Grand Challenges of Robotics [25]. Our work here on transparent interaction works towards facilitating Theory of Mind in the long term.

Errors and Transparency. It has previously been hypothesised that transparency will make it easier for users to evaluate and assess the systems' actions and responses [10]. Therefore, users will be more likely to notice an error made by the robot. Additionally, the intended action by the robot is expected to be more understandable for the user [11]. When a robot makes an error, this affects the user's trust towards the system, and lowers the user's assessment of the robot's reliability in the future [24]. Transparency has been found to reduce conflict when errors occur, and makes the system regain the user's trust more rapidly [23]. Work described in Wortham and Theodorou [23] also calls attention to the possibility of low transparency levels resulting in users making incorrect assumptions of the robot's behaviour. The user could easily interpret the robot's behaviour as correct while it was in fact an error, or assume that the robot was faulty because they do not understand its actions. Well-designed transparent interaction would again help avoid this situation.

Human-Robot Trust. Trust is a complex multifaceted phenomena that has had many definitions. We take trust to be defined as in Rousseau et al. [17] as the willingness to accept one's vulnerability for relying on the behaviour of another, if the risks and the uncertainty that this interdependence imply are counterbalanced by the positive expectations on their intentions and actions. This definition is important in our scenario as real users/patients in this scenario may be feeling vulnerable because their health is at risk, as well as the fact that transparency can reflect the user's ability to monitor/control the robot. Elements of trust are captured in Schaefer's questionnaire on Human-Robot Trust [18]. However, for this study, we chose to use the Jian Trust Survey [9], as it has been widely used in HRI and is relatively succinct with only 12 items compared to the full 40 item Schaefer scale.

3 EXPERIMENTAL DESIGN AND SETUP

We formulated three broad research questions:

- **RQ 1:** How will different levels of transparency affect the user's trust?
- **RQ 2:** How will system errors affect the user's trust, across different levels of transparency?
- **RQ 3:** Does the transparency level affect the user's ability to make good decisions?



Figure 1: Example frame from one of the video vignettes. Pepper's tablet reads: Say "Hello" to start the consultation.

3.1 Conditions

The study was done through an on-line survey with a 2x2 within-subject (or repeated measures) design. Such design was chosen to minimise the random noise and confounding variables (e.g. different times of the day). In addition, within-subject studies also tend to reduce the variability between subjects, which is key for trust studies where people vary in both their propensity to trust and their attitude to robots [6, 8, 13]. Two independent variables were considered: Transparency and Error, each manipulated on two levels. Thus, there were four conditions: Low Transparency No Error (LT), High Transparency No Error (HT), Low Transparency With Error (LTE), High Transparency With Error (HTE).

Each participant watched four video vignettes, one for each condition. Each presented a diagnostic consultation between a high risk patient with 60 years of age suffering from asthma and a Pepper robot acting as a COVID-19 triage expert system. Each video ends with the patient either being diagnosed as needing a COVID-19 test or being advised they need no test, with the detailed content of the diagnosis statement varying with experiment condition. Following each video, participants completed self-report questions presenting the measurements: (1) Trust from a twelve item Trust Survey reported in [9] and (2) Likelihood to request a second opinion (2ndOp), measured by a single 7-point Likert item.

We used the 2ndOp measure to gauge the effect of the Transparency and Error conditions on participants' ability to make good decisions based on perceptions of the interaction with the robot.

3.2 The Video Vignettes

Each video (see Figure 1) showed a male able to pass as 60 years old in a consultation setting with a Pepper robot. The text on the Pepper robot's tablet was clearly shown and subtitles were included to ensure that participants fully understood the dialogue.

The volunteer, posing as the patient, followed a script for each of the four consultations. The Pepper robot responded as programmed by a dialogue manager created using the *Choregraphe* development tool. The video scripts were developed together with a subject matter expert, a Senior Consultant actively involved in diagnosing COVID-19 patients at a major European University Hospital. The questions asked by the robot address the following topics:

- Symptom diagnosis.
- Gathering of personal details, i.e. name, date of birth and gender (important in establishing individual risk factors).

- Establishing a pattern of potential contact of people tested positive for COVID-19.

The composition of these four video conditions is set out below. First, we make clear the rationale for the two levels of the Error condition, and then we specifically list the robot advice for all four conditions.

- **No Error condition** - Patient presents with symptoms, is diagnosed correctly by the robot and is requested to follow the robot to be taken for testing.
- **With Error condition** - Patient presents with symptoms, the robot makes an error and sends the patient home.

The consultation ends with the robot's advice and it is this that varies according to the experimental conditions as described here:

- **Low Transparency No Error (LT):** *"Noted. Please follow me". [Robot leads patient to be tested.]*
- **High Transparency No Error (HT)¹:** *"Noted, due to your asthma and current conditions, there is a high likelihood that you might have Corona. I think we should have you tested for the virus. Please follow me into the examination room so we can take some tests". [Robot leads patient to be tested.]*
- **Low Transparency With Error (LTE):** *"Noted, there is no need for you to isolate or take any other safety precautions. I don't think you should get tested for the virus, and you are free to go home". [Robot does not move.]*
- **High Transparency With Error (HTE):** *"Noted, there seems to be nothing wrong with you. Since you have no symptoms related to the coronavirus, there is no need for you to isolate or take any other safety precautions. I don't think you should get tested for the virus, and you are free to go home". [Robot does not move.]*

Embodiment, voice and gestures remain the same for all conditions. In some conditions the robot requires the patient to physically follow it, thus justifying its use over a static agent.

3.3 Hypotheses

We formed the following hypotheses:

- **H1:** There is a difference in the user's trust between the high and low transparency, when the robot makes no error.
- **H2:** There is a difference in the user's trust between the high and low transparency, when the robot makes an error.
- **H3:** There is a difference in the participants' likelihood to request a second opinion between the high and low transparency, when the robot makes no error.
- **H4:** There is a difference in the participants' likelihood to request a second opinion between the high and low transparency, when the robot makes an error.

3.4 Participants and recruitment

Participants were recruited through convenience sampling, email and social media. 61 people took part. 3 sets of responses were rejected as they failed an attention check. Of the remaining 58 participants who completed the study, 31 identified as male, 26 as female, and one preferred not to disclose their gender. The average age of the participants was 33.4 years ($SD = 11.9$ years).

¹Link to HT video as example <https://www.youtube.com/watch?v=dpKUoQeRYYI>

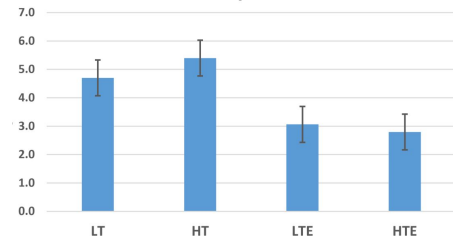


Figure 2: Trust score means for the four conditions. Error bars show standard deviations.

3.5 Procedure

The questions were presented in an online questionnaire consisting of: an introduction, a consent form, and a pre-experiment demographics questionnaire; secondly, the four conditions with accompanying subjective measure questionnaires, and; lastly an opportunity to give any feedback after the experiment.

When undertaking the four conditions, participants viewed all four videos in a fixed order and after each completed the two self-reported measures. The experimental design decision was made in both scenarios (No Error/With Error) to show the low transparency first to avoid the learning effect. Showing the high transparency scenario first, would likely introduce a bias, given the robot revealed details of its internal processing behind the decision making, which would likely influence the participant's perception of the robot. The order of the No Error/With Error conditions was also fixed for similar reasons, i.e. if the error-full robot was shown initially, this would influence their perception of the error-free robot and potentially could lower the overall trust in such a way that alternate or random ordering would not provide a fair comparison between the conditions. As a further step to try to avoid the impression of the robot in Scenario 1 affecting those of Scenario 2, we mentioned to the subjects that the robot had been "reprogrammed". This was so as to try to limit the effect of the perceived reliability/trustworthiness of the robot already built up in the high transparency condition for Scenario 1, where the participant already has some pre-conceived ideas of the robot's reasoning due to the transparent nature of the interaction. The ordering is thus: (1) Low Transparency No Error (LT), (2) High Transparency No Error (HT), (3) Low Transparency With Error (LTE) and (4) High Transparency With Error (HTE).

4 RESULTS AND ANALYSIS

In this section, we report our results and statistical analysis. We used Two-Way Repeated Measures ANOVA for our Trust scale data and Friedman's Two-Way Analysis of Variance by Ranks for our opinion rating data, both carried out in IBM SPSS. We used ANOVA on the trust scale data as the scale produces continuous data and we use Friedman's (a non-parametric test) on our Likert data, which is ordinal [5]. Results are documented below, in Figures 2 and 3, and Tables 1, 2 and 3. The data will be made available on our institutional repository following publication.

4.1 Hypotheses Revisited

H1 - We reject the null hypothesis here. There is a statistically significant difference in LT-Trust vs HT-Trust. See Table 2, row 1

Table 1: Trust measure: Two-way Repeated Measures ANOVA.

Effect source	F-Statistic	p-value
Transparency	$F(1)=146.81$.021
Error	$F(1)=5.68$	<.001
Transparency*Error (interaction)	$F(1)=16.31$	<.001

Table 2: Trust measure: post hoc Pairwise Comparisons by Paired samples t-Tests with Bonferroni correction, to isolate the simple main effects of Transparency and Error.

Row No.	Pair	t-Statistic	p-value (Bonferroni Adjusted)
Transparency			
1	LT-Trust vs HT-Trust	$t(57)=-4.519$	<.001
2	LTE-Trust vs HTE-Trust	$t(57)=1.830$.144
Error			
3	LT-Trust vs LTE-Trust	$t(57)=8.807$	<.001
4	HT-Trust vs HTE-Trust	$t(57)=11.078$	<.001

Table 3: The Likelihood to Seek a Second Opinion (2ndOp) measure post hoc pairwise comparisons by Dunn-Bonferroni tests.

Row No.	Pair	Z-Statistic	p-value 2-tailed (Bonferroni Adjusted)
1	LT-2ndOp - HT-2ndOp	-2.805	.030
2	HTE-2ndOp - LTE-2ndOp	2.230	.155
3	LTE-2ndOp - HT-2ndOp	5.825	<.001
4	HTE-2ndOp - LT-2ndOp	5.250	<.001
5	LTE-2ndOp - LT-2ndOp	3.021	.015
6	HT-2ndOp - HTE-2ndOp	8.055	<.001

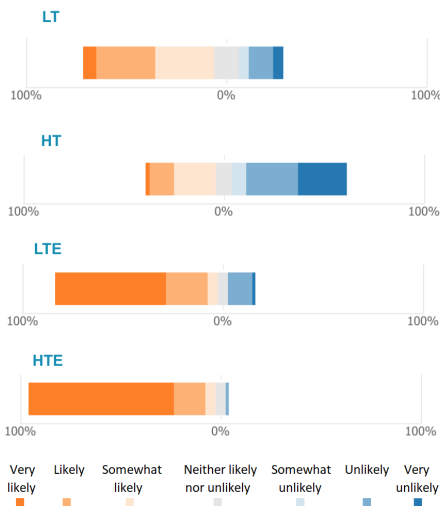


Figure 3: The relative proportions by condition of each of the responses given to the Likert item for self-report of Likelihood to request a second opinion (2ndOp). N=58.

and Figure 2. Participants trusted the robot significantly more in the high transparency interaction when the robot made no error, compared to the low transparency interaction.

H2 - Here, we accept the null hypothesis. There is no statistically significant difference in LTE-Trust vs HTE-Trust (Table 2, row 2).

H3 - We reject the null hypothesis. There is a statistically significant difference in LT-2ndOp vs HT-2ndOp ($M,3.36, Mdn,3.0$ vs $M,4.81, Mdn,5.5$). See Table 3, row 1. Participants were more likely to request a second opinion in the low transparency interaction, when the robot makes no error, compared to the high transparency interaction when the robot makes no error.

H4 - Although it is not statistically significant, participants were more likely to request a second opinion in the HTE interaction, compared to LTE (HTE: $M,1.41, Mdn,1.0$ vs LTE: $M,2.47, Mdn,2.0$; where low means more likely to request a second opinion). See Figure 3.

5 CONCLUSION AND FUTURE WORK

To investigate the effect of designing transparency into a health diagnostic system embodied as a social robot, we conducted a 2x2 repeated measures experiment with two factors, Transparency and Error, each with two levels.

Our findings show that in this COVID-19 robot diagnosis scenario, designing Transparency into such a system appears to lead people to make better judgements about the system and hence better health decisions based on the robotic system’s advice. Our results were in line with previous work with respect to the fact that errors impact trust in HRI (RQ2) [24]. High transparency (HT) interactions had significantly higher trust when compared with low transparency (LT) when no error was made by the robot. When there was an error, the same trend does not hold. Therefore, transparency seems to have a positive impact on trust when the robot makes no error (RQ1). When transparency was high (HT) participants correctly calibrated their trust to be high, such that they were accepting of robot advice and were least likely (of all the conditions) to seek a second opinion (RQ3). When the system made an error and transparency was high (HTE) participants correctly calibrated their trust to be lower and were most likely (of all the conditions) to seek a second opinion. It is also important to ensure that any explanations are targeted at the right level for the user’s background knowledge and expertise [7]. For example, in our interaction this would vary depending on if the user is from the medical profession or not [14]. Our findings can be used to further the research on system transparency, particularly for social robotic health diagnostic systems and how transparency affects realistic and appropriate trust development in users.

Future work could involve replicating the results of this experiment in a first-person setting instead of doing a perception study. In addition to the two conditions studied, a third one where the robot realises the mistake and continues the interaction could be added to the experiment. We will also investigate potential correlation with dialogue length and trust in future work.

ACKNOWLEDGMENTS

We are grateful to our funder, EPSRC EP/R026173/1 ORCA Hub. This work was also supported by the UKRI Trustworthy Autonomous Systems Node in Trust (EP/V026682/1, 2020-2024).

REFERENCES

- [1] Adella Bhaskara, Michael Skinner, and Shayne Loft. 2020. Agent Transparency: A Review of Current Theory and Evidence. *IEEE Transactions on Human-Machine Systems* (2020), 1–10.
- [2] J. Bryson and A. Winfield. 2017. Standardizing Ethical Design for Artificial Intelligence and Autonomous Systems. *Computer* 50, 5 (2017), 116–119.
- [3] Amol Deshmukh, Srinivasan Janarthanam, Helen Hastie, Mei Yui Lim, Ruth Aylett, and Ginevra Castellano. 2016. How Expressiveness of a Robotic Tutor is Perceived by Children in a Learning Environment. In *Proceedings of the International Conference on Human-Robot Interaction (HRI)*.
- [4] EPSRC. 2020. Principles of robotics. <https://epsrc.ukri.org/research/ourportfolio/themes/engineering/activities/principlesofrobotics/>
- [5] Andy Field. 2009. *Discovering Statistics Using SPSS* (3rd ed.). Sage, London.
- [6] M. Lance Frazier, Paul D. Johnson, and Stav Fainshmidt. 2013. Development and validation of a propensity to trust scale. *Journal of Trust Research* 3, 2 (2013), 76–97. <https://doi.org/10.1080/21515581.2013.820026>
- [7] Francisco J. Chiyah Garcia, David A. Robb, Atanas Laskov, Xingkun Liu, Pedro Patron, and Helen Hastie. 2018. Explainable Autonomy: A Study of Explanation Styles for Building Clear Mental Models through a Multimodal Interface. In *Proceedings of the International Conference on Natural Language Generation (INLG'18)*.
- [8] Harjinder Gill, Kathleen Boies, Joan E Finegan, and Jeffrey McNally. 2005. Antecedents of trust: Establishing a boundary condition for the relation between propensity to trust and intention to trust. *Journal of business and psychology* 19, 3 (2005), 287–302.
- [9] Jiun-Yin Jian, Ann M Bisantz, and Colin G Drury. 2000. Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics* 4, 1 (2000), 53–71. http://www.tandfonline.com/doi/abs/10.1207/S15327566IJCE0401_04
- [10] Taemie Kim and P Hinds. 2006. Who Should I Blame? Effects of Autonomy and Transparency on Attributions in Human-Robot Interaction. (2006), 80–85.
- [11] T. Kulesza, S. Stumpf, M. Burnett, S. Yang, I. Kwan, and W-K Wong. 2013. Too much, too little, or just right? Ways explanations impact end users' mental models. *Proceedings of IEEE Symposium on Visual Languages and Human-Centric Computing, VL/HCC* (2013), 3 – 10. <https://doi.org/10.1109/VLHCC.2013.6645235>
- [12] Joseph B Lyons. 2013. Being transparent about transparency: A model for human-robot interaction. In *2013 Association for the Advancement of Artificial Intelligence (AAAI) Spring Symposium Series*. 48–53.
- [13] Tatsuya Nomura, Tomohiro Suzuki, Takayuki Kanda, and Kensuke Kato. 2006. Measurement of negative attitudes toward robots. *Interaction Studies* 7 (01 2006), 437–454. <https://doi.org/10.1075/is.7.3.14nom>
- [14] François Portet, Ehud Reiter, Albert Gatt, Jim Hunter, Somayajulu Sripada, Yvonne Freer, and Cindy Sykes. 2009. Automatic generation of textual summaries from neonatal intensive care data. *Artificial Intelligence* 173, 7 (2009), 789–816. <https://doi.org/10.1016/j.artint.2008.12.002>
- [15] David Premack and Guy Woodruff. 1978. Does the chimpanzee have a theory of mind? *Behavioral and brain sciences* 1, 4 (1978), 515–526.
- [16] Sandeep Reddy, Sonia Allan, Simon Coghlan, and Paul Cooper. 2019. A governance model for the application of AI in health care. *Journal of the American Medical Informatics Association* 27, 3 (11 2019), 491–497. <https://doi.org/10.1093/jamia/ocz192> arXiv:<https://academic.oup.com/jamia/article-pdf/27/3/491/32500029/ocz192.pdf>
- [17] Denise Rousseau, Sim Sitkin, Ronald Burt, and Colin Camerer. 1998. Not So Different After All: A Cross-discipline View of Trust. *Academy of Management Review* 23 (07 1998). <https://doi.org/10.5465/AMR.1998.926617>
- [18] Kristin E. Schaefer. 2013. *The perception and measurement of human-robot trust*. Ph.D. Dissertation.
- [19] Sonja Stange and Stefan Kopp. 2020. Effects of a Social Robot's Self-Explanations on How Humans Understand and Evaluate Its Behavior. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (Cambridge, United Kingdom) (HRI '20). Association for Computing Machinery, New York, NY, USA, 619–627. <https://doi.org/10.1145/3319502.3374802>
- [20] Kimberly E Vanderbilt, David Liu, and Gail D Heyman. 2011. The development of distrust. *Child development* 82, 5 (2011), 1372–1380.
- [21] Dieter Vanderelst and Alan Winfield. 2018. An architecture for ethical robots inspired by the simulation theory of cognition. *Cognitive Systems Research* 48 (2018), 56–66.
- [22] Samuele Vinanzi, Massimiliano Patacchiola, Antonio Chella, and Angelo Cangelosi. 2019. Would a robot trust you? Developmental robotics model of trust and theory of mind. *Philosophical Transactions of The Royal Society B Biological Sciences* 374 (03 2019). <https://doi.org/10.1098/rstb.2018.0032>
- [23] Robert H Wortham and Andreas Theodorou. 2017. Robot transparency, trust and utility. *Connection Science* 29, 3 (2017), 242–248.
- [24] J. L. Wright, J. Y. C. Chen, and S. G. Lakhmani. 2019. Agent Transparency and Reliability in Human-Robot Interaction: The Influence on User Confidence and Perceived Reliability. *IEEE Transactions on Human-Machine Systems* (2019), 1–10. <https://doi.org/10.1109/THMS.2019.2925717>
- [25] Guang-Zhong Yang, Jim Bellingham, Pierre E Dupont, Peer Fischer, Luciano Floridi, Robert Full, Neil Jacobstein, Vijay Kumar, Marcia McNutt, Robert Merrifield, Bradley J Nelson, Brian Scassellati, Mariarosaria Taddeo, Russell Taylor, Manuela Veloso, Zhong Lin Wang, and Robert Wood. 2018. The grand challenges of Science Robotics. *Science Robotics* 3, 14 (2018). <https://doi.org/10.1126/scirobotics.aar7650>
- [26] John Zimmerman. 2005. Video Sketches: Exploring pervasive computing interaction designs. *IEEE pervasive computing* 4, 4 (2005), 91–94.