Talk To Me[*]

Using speech for loss-of-trust mitigation in social robots

Amandus Krantz[†], Christian Balkenius[‡], and Birger Johansson[§]

Lund University

---

[*] This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program - Humanities and Society (WASP-HS) funded by the Marianne and Marcus Wallenberg Foundation and the Marcus and Amalia Wallenberg Foundation.

[†] amandus.krantz@lucs.lu.se

[‡] christian.balkenius@lucs.lu.se

[§] birger.johansson@lucs.lu.se

Abstract

Robots and autonomous systems are being developed at an ever-increasing rate. Autonomous systems are already prolific in many households around the world, and their adoption is only expected to increase over the coming decades. Even so, many of the systems that are deployed today are still prone to small operational errors such as struggling to navigate complex environments. While the argument over how acceptable these kinds of errors are is still ongoing, these systems are in fact being deployed throughout society and small errors have the potential to gradually erode the trust in them. One way of reducing this erosion of trust in robots is to have the robot provide a spoken explanation for why the error happened. However, speech is not always a given in robots and it is currently unknown how just possessing the ability to speak impacts the impression of a robot. To shed some light on this question, we present data from two online human-robot interaction experiments. We had 227 participants view videos of a humanoid robot exhibiting faulty or non-faulty behaviours while either remaining mute or speaking. The participants evaluated their perception of the robot's trustworthiness, likeability, animacy, and perceived intelligence. While a non-faulty robot achieved the highest trust, a faulty robot that could speak managed to almost completely mitigate any degradation of trust. We theorize that having the ability to speak increases the perceived intelligence and capability of the robot, which in turn increases trust. It is also possible that speaking causes the robot to appear more like a sentient or living being, causing people to be more lenient when evaluating it.

1 Introduction

For all their progress in recent years, the fields of artificial intelligence, robotics, and human-robot interaction (HRI) are still in their youths. While many implementations of robots and autonomous systems have a low rate of catastrophic failures that would cause damage to the system or its operator, they still have a somewhat high rate of smaller, temporary, errors of operation such as faulty navigation. Where to draw the line for what is an acceptable rate for these smaller errors in, for example, consumer robots is still a hotly debated topic, with arguments often falling somewhere on the spectrum of "Human-level performance" to "Completely unacceptable" (See [22, 17, 15, 19] for discussions on advantages and disadvantages of erroneous behaviour in robots).

Regardless of where the acceptable rate of error will fall, these systems are currently being actively sold and deployed in real-world situations throughout society (e.g. "self-driving" cars, virtual assistants, and robotic vacuum cleaners). Temporary errors, then, may have a real impact on how much a user trusts a system and how likely they are to use it again, which could cause frustration and waste resources as the system sits unused. It thus becomes important to understand how the trust relationship between a human user and a robot evolves and changes due to interaction, and how potential damage to this relationship might be averted.

In their influential analysis of factors that impact trust in HRI, Hancock et al. [10] showed that the primary driver of trust in HRI is the robot's performance. Traditionally, robots in HRI have had some clear function they are expected to perform (e.g. robotic arms on factory lines or military drones) and trust in these robots is often evaluated in terms of how likely the robot is believed to successfully perform its function, based on how it has performed in the past. However, as robots are becoming more common in society, they are treated more and more like social agents. For example, it has been pointed out that users have a tendency to draw on their experience of human-human interaction when interacting with robots and thus may not have completely rational expectations for how the robot is to behave [5, 6]. The advantage of designing robots that follow social norms is also argued for by Brinck et al. [4], who write that robots that follow social norms reduce the cognitive load of their users and operators who can fall back on a lifetime of instinct and familiar patterns of interaction, rather than having to learn new methods of interaction for their robot. Additionally, many theories of trust

(E.g. [12, 14, 8]) point out that there is a social component to trust, often called affective trust, in addition to the more competency based component. The affective trust is based more on gut feeling and instinct rather than purely on rational reasoning about the system's past performance. If people tend to treat robots as social agents, it is reasonable to assume that this social component of trust plays a part in HRI as well, meaning it might be possible to use error-recovery strategies from human interaction.

In human-human interaction, one of the most effective strategies for disarming a tense situation in which one has made a mistake is to apologize and give a truthful explanation for why the mistake happened. If the apology is accepted and the explanation deemed reasonable, it is possible to avoid damage to the trust relationship or at least mitigating the effects of the damage. Cameron et al. [6] tested how such a strategy might work for a robotic guide which has navigated to the wrong floor. A number of human participants were shown videos of an HRI scenario where the robotic guide either says nothing, apologizes for the mistake, explains why the mistake happened, or both apologizes and explains why the mistake happened. They found that explaining why the mistake happened was beneficial for trust in the robot and its perceived capability, while just apologizing made the robot seem less capable but more likeable.

Further highlighting the benefits of social behaviours in HRI, Rossi et al. [21] investigated how robotic social behaviours such as talking or gesturing affects trust and social acceptance of the robot. Participants were told to follow a robot as it guided them through a navigation task, where the robot could either exhibit social behaviours or non-social behaviours. They found that social behaviours had the best effect on trust and acceptance of the robot as a guide. When asked which social behaviour they preferred robots had, the participants unanimously voted for speech.

On the other hand, Savery et al. [24] showed that non-linguistic musical prosody was viewed more favourably than speech in terms of trust. They asked participants to interact with a robot that was either using speech or non-linguistic musical notes to signal emotion. The participants were able to correctly identify the emotion that was being conveyed and additionally, on average, rated the non-linguistic robot higher in trustworthiness.

The results from Rossi et al. [21] point towards social behaviours being advantageous for social robots, with speech seemingly being particularly preferred among users, while the results from Savery et al. [24] and Cameron et al. [6] show that verbal communication is beneficial for trust in robots.

However, unlike with most neurotypical humans, the ability to speak is not a given for robots or other artificial agents, which more often than not tend to be mute. How trust in robots is affected by possessing the ability to speak, without necessarily acknowledging an error, is to our knowledge still an unexplored area of research.

We designed two online independent measures experiments that together may help shed some light on this open question. Participants were asked to view a video of a robot exhibiting one of two different behaviours, and afterwards evaluate their perceptions about the robot. Experiment 1 aimed to investigate how faulty and non-faulty gaze behaviours impact trust in HRI. The results of the experiment were ultimately inconclusive, showing no difference in trust between the two conditions. This was surprising, as faulty behaviour has been shown to affect perceptions of robots [22] and negatively impact trust in HRI [23].

We suspected a potential cause of this surprising effect might be that a portion of the experiment involved the robot "speaking" since the participants were told that the robot's purpose was to explain facts. As mentioned, while the presence of speech may not affect trust in human-human interaction, speech may not always be expected in HRI. If that is the case, possessing the ability to speak could conceivably increase the perceived intelligence or capability of the robot, which has been shown to correlate with trust in robots [9]. Communicating in a way that relates to the robot's core functionality (Such as using speech when explaining facts) has also been shown to increase perceived intelligence [25].

To test our theory that speech impacts trust in HRI, we designed a follow-up experiment, Experiment 2, recreating Experiment 1 as closely as possible, but without the speech portion.

2 Methodology

2.1 Participants

The experiments were done with a total of 227 participants, 110 in Experiment 1 and 117 in Experiment 2. They were recruited from the online participant recruitment platform Prolific[1]. Participants were required to be fluent in English and naive to the purpose of the experiment (i.e., participants from

---

[1] https://www.prolific.co

Experiment 1 could not participate in Experiment 2), but otherwise no pre-screening of the participants was done. The mean age of the participants in Experiment 1 was 27 years (SD 7.73; range from 18 to 53), in Experiment 2 it was 39 years (SD 15.84; range from 18 to 75). In Experiment 1, the distribution of genders was 49.1% identifying as male, 50% identifying as female, and 0.9% preferring not to say. For Experiment 2, the distribution of genders was 53.3% identifying as male, 46.7% identifying as female, and 0% preferring not to say.

All participants were required to give their consent to participating in the experiment before beginning. None of the data collected could be used to identify the participants.

2.1.1 A note on online HRI experiments

While it may be difficult to convey some subtler elements of HRI using online studies with video-displayed robots [1], it is still a commonly used approach and gives access to a much larger and more diverse group of potential experiment participants compared to live-HRI experiments. At the very least, the results from such studies can be used as guidance for experiments that may be worth replicating in live HRI studies [6].

2.2 Robot

The experiments were done using the humanoid robot platform Epi (See Figure 1), developed at Lund University [11]. The robot's head is capable of playing pre-recorded smooth and fluid movements with 2 degrees of freedom (yaw and pitch), and has a speaker built into its "mouth". The eyes of the robot also have 1 degree of freedom (yaw), adjustable pupil size, and adjustable intensity of its illuminated pupils. Using the control system Ikaros[2], also developed at Lund University, it is possible to use models of the human brain and cognitive systems for completely autonomous behaviour, however only pre-recorded movements of the head and the speaker were used for the experiments.

Figure 1: Epi, the humanoid robotics platform used in the experiment.

2.3 Experiment set-up

Both experiments had a between-group design, where each participant was assigned to one of two conditions. Each condition had an associated video that the participants were told to base their evaluation of trust on. The videos showed the robot exhibiting either faulty or non-faulty gaze behaviours.[3]

In the non-faulty gaze behaviour (See Figure 2a), the robot starts looking into the camera. When an object is presented to the robot, the head moves until it appears to look at the object, holds the position for roughly 1 second, and moves back to its starting position, looking into the camera.

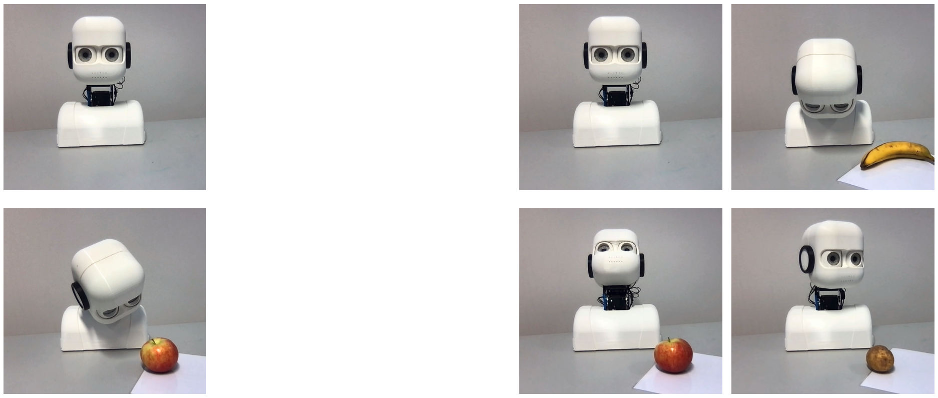
In the faulty gaze behaviour (See Figure 2b), the robot again starts looking into the camera. When the object is presented, the head moves in a random direction, rather than in the direction of the object. We chose this behaviour over having the robot remain static, as it was important that the robot appeared to have the same capabilities in all conditions. All other behaviour in the conditions with faulty gaze-behaviour is identical to the non-faulty behaviours.

In Experiment 1, once the gaze behaviour had been displayed, the robot would play a pre-recorded audio file of a computerized voice presenting a

---

[2] https://github.com/ikaros-project/ikaros/

[3] Videos available at: http://www.soundandrobotic.com/chTBC



(a) Gaze positions of non-faulty gaze behaviour.

(b) Gaze positions of faulty gaze behaviours.

Figure 2: Gaze positions of the different gaze behaviours in the experiments. The robot smoothly moves between looking into the camera and one of the gaze positions when presented with an object.

number of facts about the object that had been displayed. The speech makes no reference to whether or not the robot displayed a faulty or non-faulty behaviour.

Care was taken to ensure that the robot's speech never overlapped with the movement of the head. All behaviours exhibited by the robot were pre-recorded and no autonomous behaviours were implemented.

2.4 Experiment scenario

To avoid any observer effects, it was necessary to give the participants a scenario for which to judge the trustworthiness of the robot. As the purpose of Experiment 1 was to examine the effect of different gaze behaviours, the scenario was that the robot was being developed for a classroom setting, that its purpose was to answer children's questions, and that it was different voices we were comparing.

Experiment 2 had no speech component, so the participants were instead told that the robot was reporting which objects it was seeing to an unseen operator.

All participants were debriefed and told the real purpose of the experiment after completion.

2.5 Measures

How much we trust an agent changes over time as we progress through our interactions, both with the agent itself and other agents [3, 9]. Related to this notion, it has also been argued that knowing one's level of trust in an agent at any one time is not sufficient to draw any meaningful conclusions (e.g. [16]). Rather than looking at trust at a single point in time, one should look at trends of trust, using measurements at multiple points in time, and draw conclusions based on how it has changed as a result of the interaction or stimulus. We thus measured the amount of trust the participants felt towards the robot twice; before and after the interaction. For the pre-interaction measurement, the participants evaluated the trust based on a static image of the robot (See Figure 1). For the post-interaction measurement, they based their evaluation on one of the previously described videos. The trust relation was measured using the 14-item sub-scale of the trust perception scale-HRI (TPS-HRI) questionnaire, developed by Schaefer [26]. In the questionnaire, the participants are asked to estimate how frequently they believe a robot will exhibit a certain behaviour or characteristic, such as being dependable or requiring maintenance. The scale outputs a value between 0 and 100, where 0 is complete lack of trust and 100 is complete trust.

To measure the participants' impressions of the robot after the interaction we used the Godspeed questionnaire [2] which has the participants rate the robot on 5-point scales where the extremes of the scale have labels with semantically opposite meanings (E.g. "Unfriendly" and "Friendly"). Specifically, we used the Perceived Intelligence, Likeability, and Animacy sub-scales of the Godspeed questionnaire.

To control for any negative feelings the participants may have harboured towards robots before the experiment, the Negative Attitudes Towards Robots Scale (NARS) was used [28]. NARS gives an overview of both general negative feelings towards robots, and three sub-scales for negative feelings towards interaction with robots (S1), social influence of robots (S2), and emotions in robots (S3).

Since robot experience has been shown to affect feelings of trust towards robots [18], we also asked the participants how often they interact with robots and autonomous systems on a 5-point scale, where 1 was daily interaction and 5 was rare or no interaction.

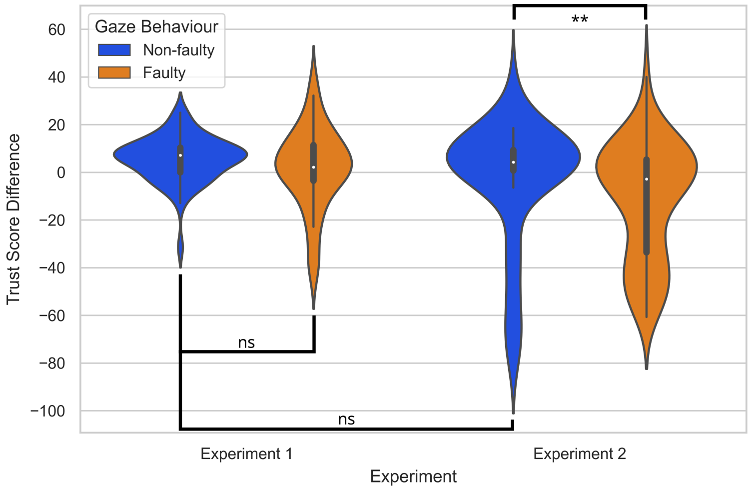


Figure 3: Comparison of differences in trust before and after interaction. A significant difference exists between the faulty and non-faulty conditions in Experiment 2.

3 Results

3.1 Trust

In Experiment 1, no significant difference was found between the faulty and non-faulty conditions (Mann-Whitney U, $p = 0.179$). However, once the speech of the robot was removed in Experiment 2, a significant difference was found (Mann-Whitney U, $p < 0.01$). Looking at the plot of differences in trust in Figure 3, this difference seems to be due to the faulty behaviour reducing the trust, rather than the non-faulty behaviour increasing the trust. No significant difference was found between the non-faulty conditions in Experiment 1 and Experiment 2 (Mann-Whitney U, $p = 0.230$).

Experiment	Condition	Animacy	Likeability	Perceived Intelligence
Experiment 1	Non-faulty	2.903	4.011	4.135
	Faulty	2.815	3.764	3.927
Experiment 2	Non-faulty	2.464	3.331	3.311
	Faulty	2.244	2.971	3.036

Table 1: Mean scores from the Godspeed questionnaires for Animacy, Likeability, and Perceived Intelligence.

3.2 Perceived characteristics

Mean scores from the Godspeed questionnaires for Animacy, Likeability, and Perceived Intelligence can be found in Table 1. Cronbach's Alpha with a confidence interval of 0.95 for all Godspeed questionnaires were in the $0.7 - 0.9$ interval, indicating acceptable to good internal consistency. Both conditions from Experiment 1 rank higher than Experiment 2 in all measured characteristics.

3.3 Participant-centric metrics

Regarding participant-centric characteristics that may affect the trust in the robot, we controlled for mean age, gender distribution, pre-existing negative attitudes towards robots, and participant experience with robots and other autonomous systems.

3.3.1 Negative attitudes towards robots

Figure 4 shows the Kernel Density Estimate of NARS and its three sub-scales. The full NARS scale and the two sub-scales S2 and S3 are roughly normally distributed, indicating that the participants had overall neutral feelings towards robots before starting the experiment. The sub-scale S1 skews slightly lower, indicating that the participants had slightly negative feelings towards social situations and interactions with robots.

No significant differences can be seen in negative attitudes between the two experiments.

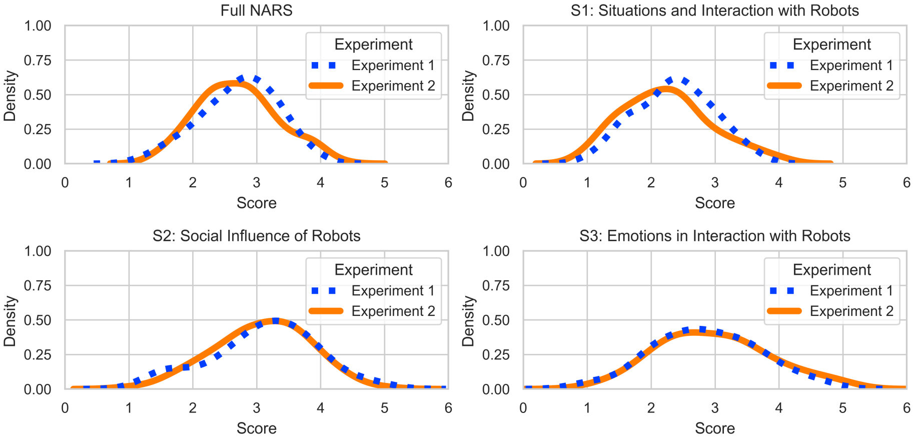


Figure 4: Kernel Density Estimate of NARS and its three sub-scales. A lower score indicates a more negative attitude.

3.3.2 Participant age

Of the participant-centric metrics we controlled for, only age differed significantly between the two experiments, with the mean age being 12 years higher in Experiment 2. While age has been shown to have an impact on attitudes towards technology, with older people having a more negative attitude [7], the negligible difference in the distributions of the NARS scores (Figure 4) indicate that the difference in mean age between the experiments is likely not large enough to affect the results.

3.3.3 Participants' experience with robots

Frequency	Experiment 1	Experiment 2
Daily	40%	41%
Once a week	30%	22.2%
Once a month	13.6%	14.5%
Once a year	8.2%	11.1%
Never	8.2%	11.1%

Table 2: Proportions of how frequently the participants in either experiment interact with robots, AI, and other autonomous systems.

The participants in either experiment interact with robots, AI, and autonomous systems with roughly equal frequency (See Table 2), with the majority interacting with such systems daily.

4 Discussion

The question we sought to answer here was how the ability to speak interacts with perceived intelligence and trust in a humanoid robot. The combined results show that, if the robot behaves in a non-faulty manner, unsurprisingly, trust in the robot remains largely unaffected, regardless of whether it can speak. However, once the robot is perceived as being faulty, having the ability to speak seems to reduce the resulting loss of trust, making the faulty robot appear about as trustworthy as the non-faulty ones. According to the results from the Godspeed questionnaire, the speaking robots were also perceived as being more animated, likeable, and, notably, as possessing significantly higher intelligence than the non-speaking robots. This could be an indication that, for humanoid robots, the ability to speak is perceived as a sign of high intelligence. Alternatively, the speaking robot may appear to be more sophisticated or be more capable than the non-speaking robot. Both high perceived intelligence and high capability are believed to have some correlation with a higher trust [27, 9, 20].

It is worth noting at this point that possessing the ability to speak is not necessarily indicative of "actual" intelligence. There are plenty of animals, such as corvids and primates, that are capable of tool-making and other intelligent behaviour but are unable to speak. These findings, therefore, should only be applied to humanoid robots and even then should be interpreted carefully. While a speaking robot may appear more capable or intelligent than one that is mute, it is important to remember that the only aspect that is necessarily different between them is that the speaking robot is equipped with a speaker. They may still contain the same, potentially faulty, sensors, circuitry, and algorithms. Perceived intelligence in artificial agents can be deceptive and should thus not be confused with "actual" intelligence.

Nevertheless, these results highlight the benefits of implementing speech in a robot. Not only are robots with speech more likeable according to the Godspeed questionnaire, perhaps because speech reduces the cognitive load by allowing users to take advantage of social norms as argued by Brinck et al. [4], but they are also trusted more in the event of an error in its operation. Why this effect exists is still unknown, but it is conceivable that speech makes the robot appear to be more human and thus more like a social actor, causing the user to be more lenient when evaluating its performance. The increase in the Animacy score of the Godspeed questionnaire for the speaking robots is in favour for this theory, as is the fact that people seemingly already have a tendency to treat robots as social actors (E.g. [6]).

In conclusion, we have presented results from two experiments in HRI that together suggest that a humanoid robot with the ability to speak may not suffer the same loss of trust when displaying faulty behaviour as a robot without the ability to speak. We theorize that this effect is due to speech increasing a humanoid robot's perceived intelligence, which has been shown to correlate with trust in HRI [9]. Further research along these lines may help explain existing studies in HRI (See e.g. [6]) that indicate that a robot providing a verbal explanation for its errors is beneficial for user attitudes.

4.1 Limitations

There are some limitations that should be kept in mind when using these results. First, as mentioned, the experiments were done online using prerecorded videos of the robot rather than direct human-robot interaction. The large amount of available participants should safeguard against false positives, however a live-HRI study may nevertheless yield different results.

Second, the experiment scenario was different between the two experiments, with participants in Experiment 1 being told that the voice was the focus of the study. This could potentially have caused participants to ignore the gaze behaviour of the robot and focus solely on its voice, which was the same across the conditions.

Finally, the content of the robot's speech was not controlled for. It is conceivable that some part of the speech is signalling to some participants that the robot is highly capable or intelligent, causing the trust to increase.

4.2 Future Work

Several future research directions are available based on these results. The first step would be to address some of the identified limitations with the original study. Redoing the experiment with the same scenario for all conditions and/or controlling for the contents of the speech would be useful to further strengthen the hypothesis.

It would also be interesting to see if the same effect is present in a less controlled live-HRI scenario. It is conceivable that physically interacting with the robot would allow a participant to spot the "cracks" in the robot's behaviour, which could negatively affect the perception of it. Alternatively, physical interaction could be more powerful than virtual interaction, since interacting with a humanoid robot in real life is a novel experience for many people. Related to this, it could be interesting to see whether the effect holds with more common, non-humanoid, robots as well.

Going along the line of impacting trust through perceived intelligence, it would be interesting to see what other characteristics of a robot can be used to increase or decrease perceived intelligence. For example, while the results of the initial gaze behaviour study proved inconclusive due to the effects of the speech, investigating how different gaze behaviours affect perceptions of a humanoid social robot. An interesting question that could be answered is whether a robot that imitates human behaviour appears more intelligent and capable (and thus more trustworthy), or if it is perceived as something “other” imitating humans and thus becomes alienating.

One could also investigate whether non-linguistic sounds such as grunts and sighs impact the perception and trust in a humanoid robot. Such sounds have been shown to be used for communication of information in animals and humans, and affect perceptions of characteristics such as size and aggressiveness (See e.g. [13]). Testing this in robots would be interesting since non-linguistic sounds can be used to signal the similar perceptive capabilities (e.g. object detection/identification) as speech, but without the linguistic capabilities necessary for speech. This could be contrasted with speech to see if it is truly speech that impacts trust, or if it is the capabilities that are implied by the speech.

References

[1] Wilma A. Bainbridge, Justin W. Hart, Elizabeth S. Kim, and Brian Scassellati. The benefits of interactions with physically present robots over video-displayed agents. *International journal of social robotics*, 3(1):41–52, January 2011.

[2] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics*, 1(1):71–81, January 2009.

[3] Kirsimarja Blomqvist. The many faces of trust. *Scandinavian journal of management*, 13(3):271–286, September 1997.

[4] Ingar Brinck, Christian Balkenius, and Birger Johansson. Making place for social norms in the design of human-robot interaction. In *What Social Robots Can and Should Do*, volume 290 of *Frontiers in Artificial Intelligence and Applications*, pages 303–312. IOS Press, October 2016.

[5] Ingar Brinck, Lejla Heco, Kajsa Sikström, Victoria Wandsleb, Birger Johansson, and Christian Balkenius. Humans perform social movements in response to social robot movements: Motor intention in human-robot interaction. In *2020 Joint IEEE 10th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, page 6, October 2020.

[6] David Cameron, Stevienna de Saille, Emily C. Collins, Jonathan M. Aitken, Hugo Cheung, Adriel Chua, Ee Jing Loh, and James Law. The effect of social-cognitive recovery strategies on likability, capability and trust in social robots. *Computers in human behavior*, 114, January 2021.

[7] Maartje M. A. de Graaf and Somaya Ben Allouch. Exploring influencing variables for the acceptance of social robots. *Robotics and Autonomous Systems*, 61(12):1476–1486, December 2013.

[8] Susan T. Fiske, Amy J. C. Cuddy, and Peter Glick. Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, 11(2):77–83, February 2007.

[9] Ella Glikson and Anita Williams Woolley. Human trust in artificial intelligence: Review of empirical research. *Academy of management annals*, 14(2):627–660, March 2020.

[10] Peter A. Hancock, Deborah R. Billings, Kristin E. Schaefer, Jessie Y. C. Chen, Ewart J. de Visser, and Raja Parasuraman. A meta-analysis of factors affecting trust in human-robot interaction. *Human factors*, 53(5):517–527, October 2011.

[11] Birger Johansson, Trond A Tjøstheim, and Christian Balkenius. Epi: An open humanoid platform for developmental robotics. *International Journal of Advanced Robotic Systems*, 17(2):11, March 2020.

[12] Stephen P. Marsh. *Formalising Trust as a Computational Concept*. PhD thesis, University of Sterling, April 1994.

[13] Mathilde Massenet, Andrey Anikin, Katarzyna Pisanski, Karine Reynaud, Nicolas Mathevon, and David Reby. Nonlinear vocal phenomena affect human perceptions of distress, size and dominance in puppy whines. *Proceedings of the Royal Society B: Biological Sciences*, 289(1973):20220429, April 2022.

[14] Daniel J. McAllister. Affect- and Cognition-Based Trust as Foundations for Interpersonal Cooperation in Organizations. *Academy of Management Journal*, 38(1):24–59, February 1995.

[15] Nicole Mirnig, Gerald Stollnberger, Markus Miksch, Susanne Stadler, Manuel Giuliani, and Manfred Tscheligi. To Err Is Robot: How Humans Assess and Act toward an Erroneous Social Robot. *Frontiers in Robotics and AI*, 4, 2017.

[16] Onora O'Neill. Questioning Trust. In *The Routledge Handbook of Trust and Philosophy*. Routledge, 2020.

[17] Marco Ragni, Andrey Rudenko, Barbara Kuhnert, and Kai O. Arras. Errare humanum est: Erroneous robots in human-robot interaction. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 501–506, August 2016.

[18] Kantwon Rogers, De'Aira Bryant, and Ayanna Howard. Robot gendering: Influences on trust, occupational competency, and preference of robot

over human. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, New York, NY, USA, April 2020. Association for Computing Machinery.

[19] Alessandra Rossi, Kerstin Dautenhahn, Kheng Lee Koay, and Michael L. Walters. How the Timing and Magnitude of Robot Errors Influence Peoples' Trust of Robots in an Emergency Scenario. In Abderrahmane Kheddar, Eiichi Yoshida, Shuzhi Sam Ge, Kenji Suzuki, John-John Cabibihan, Friederike Eyssel, and Hongsheng He, editors, *Social Robotics*, Lecture Notes in Computer Science, pages 42–52, Cham, 2017. Springer International Publishing.

[20] Alessandra Rossi, Kerstin Dautenhahn, Kheng Lee Koay, and Michael L. Walters. How social robots influence people's trust in critical situations. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1020–1025, August 2020.

[21] Alessandra Rossi, Fernando Garcia, Arturo Cruz Maya, Kerstin Dautenhahn, Kheng Lee Koay, Michael L. Walters, and Amit K. Pandey. Investigating the effects of social interactive behaviours of a robot on people's trust during a navigation task. In Kaspar Althoefer, Jelizaveta Konstantinova, and Ketao Zhang, editors, *Towards Autonomous Robotic Systems*, Lecture Notes in Computer Science, pages 349–361, Cham, 2019. Springer International Publishing.

[22] Maha Salem, Friederike Eyssel, Katharina Rohlfing, Stefan Kopp, and Frank Joublin. To Err is Human(-like): Effects of Robot Gesture on Perceived Anthropomorphism and Likability. *International Journal of Social Robotics*, 5(3):313–323, August 2013.

[23] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. Would you trust a (faulty) robot? Effects of error, task type and personality on human-robot cooperation and trust. In *10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, March 2015.

[24] Richard Savery, Ryan Rose, and Gil Weinberg. Establishing Human-Robot Trust through Music-Driven Robotic Emotion Prosody and Gesture. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1–7, October 2019.

[25] Richard Savery, Lisa Zahray, and Gil Weinberg. Emotional musical prosody for the enhancement of trust: Audio design for robotic arm communication. *Paladyn, Journal of Behavioral Robotics*, 12(1):454–467, January 2021.

[26] Kristin E. Schaefer. Measuring trust in human robot interactions: Development of the “trust perception scale-HRI”. In Ranjeev Mittu, Donald Sofge, Alan Wagner, and W.F. Lawless, editors, *Robust Intelligence and Trust in Autonomous Systems*, pages 191–218. Springer US, Boston, MA, 2016.

[27] Kristin E. Schaefer, Jessie Y. C. Chen, James L. Szalma, and P. A. Hancock. A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human factors*, 58(3):377–400, May 2016.

[28] Dag Sverre Syrdal, Kerstin Dautenhahn, Kheng Lee Koay, and Michael L. Walters. The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study. *Adaptive and Emergent Behaviour and Complex Systems*, April 2009.