A Human-Centered Approach to Evaluating Robot Theory of Mind (RToM)

Kerstin Haring
University of Denver
Department of Computer Science
Denver, CO
https://orcid.org/0000-0001-6338-5575

Dan Pittman
University of Denver
Department of Computer Science
Denver, CO
https://orcid.org/0000-0002-0888-3587

Benjamin Dossett
University of Denver
Department of Computer Science
Denver, CO
https://orcid.org/0000-0003-4831-1943

***Abstract*—While current research indicates that humans might form a Theory of Mind (ToM) while interacting with a social robot, existing research methods in Human-Robot Interaction often focus on the assessment of only a small number of distinct robot designs. The lack of diversity in existing robot designs also prevents meaningful analysis into ToM using Machine Learning (ML) techniques, as there is no large enough training corpus to provide reliable results. Even if this lack of data was addressed and a ML model was created, however, there is no comprehensive independent metric that could evaluate the correctness of such learned models, a key component in building trust in the ML predictions. To address these issues we developed a novel research platform, implemented as a web-accessible 3D game, that will allow us to gather a large number of user-provided robot design examples that are associated with a mental state. The goal is to use this data to train machine learning models that can validate and predict the Robot Theory of Mind (RToM) associated with a given design. We will then use neuroscientific measures to independently validate the data collected from the platform as well as the machine learning models. The integration of a novel robotic design platform for data collection, ML models, and neuroscience measures will allow us to gain significantly new insights into a Robot Theory of Mind.**

I. Introduction

A theory of mind (ToM) is the cognitive capacity to attribute minds to others and describes the ability to perceive mental states in others [3]. ToM or mental state reasoning represents a critical cognitive input for behavior explanation, action prediction, and moral evaluation [49]. While implementing a theory of mind in a robot would have a tremendous impact on how robots can interact socially with people [38], research to date has only achieved potential precursors of an implementation of a Theory of Mind in robots [8]. In this work we seek to explore whether humans form a Robot Theory of Mind (RToM), utilizing novel methods and technologies to answer this question.

This work takes a foundational approach and focuses solely on the impact of robot morphology on forming of a Robot Theory of Mind. One of the first things humans turn to when forming a mental model of a robot are the surface-level cues they experience when first seeing a robot. Creating an unambiguous first impression could contribute to more effective communication with the robot, [33] and a clear mental model of the robot is elicited if the robots match their task [15].

Theory of Mind involves the ability to understand that other's thoughts can be different from your own [39]. It is called a theory because we cannot know with certainty what another person might be thinking. All we can rely on to infer their intentions is our theory that we developed based on their appearance, behaviors, and our knowledge. If we assume that people form a Theory of Mind of a robot, we assume that they have mental state concepts of the robot such as "believe," "know," "want," and "see" and that they use these concepts to predict and explain robot behaviors. Thus, a human with a Robot Theory of Mind (RToM) believes that mental states play a causal role in generating behavior and infers the presence of mental states in robots by observing their appearance and behaviors [22].

If people develop a RToM they would also, at a minimum, develop empathy towards robots. Empathy refers to emotional awareness of others' feelings, or in case of robots, simulated expressions of feelings. Prior research has assessed that people indeed show empathy towards robots [37, 36, 42, 44, 9] and that robot morphology impacts perceived empathy [50]. It is unclear, however, which exact robot morphology features elicit human empathy. Research has also shown that people anthropomorphize robots, especially those that show a more human-like morphology [43, 5, 14, 30, 45, 6]. Additionally, research has found that robots are perceived to "see", "want", "know" [12, 10], and to be trusted [40, 35, 34]. Further evidence that humans indeed might form a RToM is that when the tests used to evaluate a Theory of Mind for humans (white lie test, behavioral intention task, facial affect inference, vocal affect inference, and false-belief test) are applied to robots, it has been shown that people do implicitly ascribe mental states to robots [2].

It seems that robots might have all the necessary components that would lead humans to create a Robot Theory of Mind. Given this, we postulate that by evaluating enough robot designs, we can show that humans indeed form a RToM. Further, we hypothesize that we can determine which robot features de- and increase a RToM and that we are able to predict what mental states are ascribed to a certain robot morphology.

II. Background

In order to prove our hypothesis, we need to explore a wide range of robot design characteristics that entail both higher order cognition (e.g. rationality and logic) and emotion (e.g. feelings and experience [24, 47]. These two core capacities map onto the two dimensions of mind perception: agency and experience [16]. It has been demonstrated that people automatically evaluate a target's mind along these two dimensions [16], and non-human targets can be living entities such as animals [13], and non-living entities such as robots [47, 17].

Mechanical robots historically are attributed with mainly agency but not experience [25, 20, 18]. Modern social robots however have the ability to display experiences [29]. Robot morphology has been shown to generally influence human expectations of the robot [19], perceived trust [50, 26, 41], intelligence [27, 31], and warmth and competence [4, 28]. However, research so far has been inconclusive about the causal relationships between robot morphology and perceived social functions in the robot [48, 11, 21]. What most of this research agrees upon is that robot morphology matters in how we perceive robots and their abilities [19]. This led us to formulate our first hypothesis that given a certain robot morphology, people form a Robot Theory of Mind.

Traditionally, investigating the effects of robot morphology involved survey-style research in which an existing design is presented to a user, which the user evaluates. While this style of research works well for investigating individual designs, it is difficult to generalize the findings to other existing or new robot designs, due to the low number of designs evaluated.

One approach to evaluate our first hypothesis would be to assess as many existing robot design as possible one by one. However, the currently largest database of robot designs classified by their human-like appearances contains only about 250 existing robots [32]. This makes it difficult to evaluate the level of Robot ToM (RToM) that a new or existing unstudied design may elicit in a human. Instead, we reverse the research approach by utilizing target mental state attributes like agency, experience, and empathy [16, 18] and have a broad-spectrum user base create a large number of robot designs that match a given mental state attribute. With a large number of robot designs tied to a mental state attribute, we then can employ Machine Learning (ML) algorithms to evaluate the causal relationships of robot morphology and mind perception. We propose a a series of neuroscientific studies to iteratively verify the outcomes from the data collection on robot morphology and the Machine Learning models. Currently developed databases and Machine Learning models are often not verified with independent methods for their validity. Technology like Functional Near-Infrared Spectroscopy (fNIRS) has been used to understand neural responses to social robots as an implicit response evaluation [7, 23] and will be used to validate our approach while providing additional insights in novel ways of measuring interactions with social robots.

III. Experimental Design

This research addresses the first of three main challenges robot designers currently encounter: 1) A limited amount of existing robotic designs, and identifies two more challenges: 2) a substantial amount of possible robotic designs, and 3) when exploratory machine learning models are used to aid humans with those substantial design choices, they are currently not validated by independent methods.

A. *The Build-A-Bot Platform*

The Build-A-Bot platform addresses the challenge to significantly increase the amount of robot designs that then can be used for further analysis. Given our research scope, the question we seek to answer is if we can create a web-based research platform where users are given a specific mental states inspired by traditional ToM research and then design a robot accordingly.

We seek to address this question by creating a novel research platform for rapidly constructing new robot designs. Through this platform, which we call Build-A-Bot, we will engage with the community to solicit new robot designs that are tied to a specific item of RToM. This reverses the traditional evaluation approach of robot mind perception where the robot design is fixed and we ask users a number of questions about how they perceive the given design. In our new model we will present users with a platform in which they can create any robot they desire by piecing together robot parts in an interactive application.

B. *Improving Robotic Designs Using Human-Centered AI*

We identified the challenge that the Build-A-Bot platform intentionally will lead to a large database of robot designs, more than a human likely could assess in a reasonable amount of time. To build a model that learns from the user-based input of robot designs, we are looking to answer the research question if we can use the targeted ToM mental states and the associated robot designs created on the Build-A-Bot platform to train a set of Machine Learning models focused on feature extraction, validation of desired RToM attributes in proposed robotic designs, and eventually as a tool to generate new robot designs targeting a given RToM attribute.

C. *Validating RToM Perception Using Neuroscience*

An additional challenge we identified is that we are currently lacking a reliable metric to assess human-centered AI. In the way our research approach and data is structured, we will be able to use novel methods and technologies in neuroscience (i.e. fNIRS) to validate our data as well as the AI and ML models. This will allow us to answer the research question on how can a Robot Theory of Mind be measured in a reliable and efficient manner and serve as novel metric to assess the validity of the data generated from the platform and the ML models.

IV. RESULTS

In order to allow for the rapid creation of robotic designs by community participants, we created a fully functional prototype of the Build-A-Bot design platform that is publicly available online. We selected the Unity 3D game development platform [46] as the basis for Build-A-Bot, which allows our platform to be deployed on the web using WebGL. We chose to create a drag-and-drop system that allows users to combine 3D robot parts in any manner that they see fit. An example of this is shown in Figure 1.

Fig. 1. The user interface for building robots on the Build-A-Bot platform prototype. On the left, the user can choose robot parts and colors. The parts can be dragged onto the turntable platform and positioned anywhere there. At the bottom, the user can choose to scale and rotate the parts. The user also has an undo, redo, and save option for their designs at the bottom right.

Complexity of a design has been shown to influence performance on ToM tasks and to depend on frontal lobe functioning, the area that we are targeting to measure for RToM [1]. We obtain a complexity measure for a given design by assigning several attributes including organizational information such as a category and subcategory label, as well as a complexity score to each new robot part created for the platform. The complexity score for individual parts allows us to calculate an overall complexity score for a robot design.

When a user is asked to create a new robotic design we present them with a target ToM attribute to represent with their design and a set of requirements to conform to, collectively referred to as a *Scenario*. Each robot design created through our platform is tied to a single Scenario, which can then be used to group robots by the ToM attribute that they are designed for. Figure 2 below shows an example of such a Scenario. For a scenario like this example, we postulated that a user may want to use make changes to the original shape and size of our 3D robot parts. In order to allow this, we decided to add a scaling and rotation system for all robot part as seen in Figure 3.

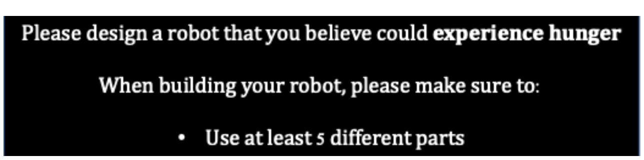

Fig. 2. An example of a scenario presented to a user before they start creating a a new robot design.

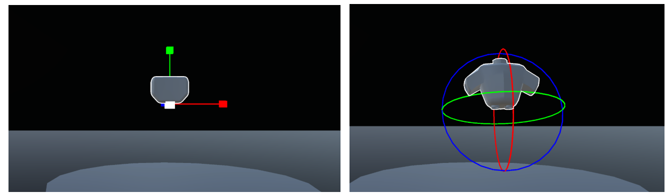

Fig. 3. Examples of the scale and rotation handles for modifying existing parts.

To broaden access of the robot platform to users that are not necessarily experienced with such an interface, we created a a tutorial that walks every first time user through the different controls and options present in the platform. Users are presented with small prompts to try out different parts of the platform while building a design. The prompts are designed to be non-intrusive for users who would rather explore the platform's features on their own, but is comprehensive enough for inexperienced users to be able to focus on what they want to design, rather than the navigation of the platform itself.

With the platform, an effective way to manage the data collected is needed. To achieve this, we are creating a researcher dashboard through our website that our team can use to perform queries against our database of submitted robots. This will allow us to look for patterns within scenarios presented to the user. Additionally, the dashboard will allow us to incrementally add and remove scenarios that will be presented to users to broaden data collection.

V. DISCUSSION

We created a fully functional web-based platform to collect robot designs that correspond to mental states that play a causal role for RToM [22]. To evaluate RToM based on the user input of robot designs, we need to have as much information as possible on the user's design choices. This will be achieved by tracking a variety of analytic measurements while the user is building their design via the Build-A-Bot platform. As the user builds a robot design in the platform, we incrementally save the changes they have made, including things like a change to the position, rotation, scale, or color of a part.

Tracking the parts used and the order in which they were chosen is important for us to be able to identify which components are important for users in the context of the ToM attribute presented to them. For example, if we find that a certain part is frequently used in designs targeting a certain ToM attribute, this would be a strong indicator that that part was important in the user creating a RToM. We are also interested in the first part a user selects, as this might indicate a key aspect to how they associate the targeted ToM attribute to their design. Once the user has indicated that their design is complete, they are no longer able to edit it, as we want a snapshot of that design that the user believed best represented the target attribute. In addition to data about individual parts, we also collect data on other aspects of the platform, such as the time the user took to create the design and which part of the platform's soundtrack was playing during the user's build session.

After the user has completed their robot design, a screenshot of the design is taken and stored. This screenshot provides a simple visual representation of the design that can be quickly re-visited during analysis or presented to users on our website. This screenshot will be used as an instrument in the other aims of our work, both for our neuroscience work and in our Machine Learning models. In the context of neuroscience, tying a screenshot of a design to the other data we collect, such as the target ToM attribute will be key in analyzing human reactions to different designs. Additionally, the screenshot may be used in our machine learning models when performing feature selection.

The other data we collect through the Build-A-Bot platform, such as the JSON data representing the robot and the path taken to create it, will also be instrumental in the other aims of our work. We plan to use this data to build ML models that allow us to create new designs targeting a specified RToM attribute. ML models like these require a significant amount of data, which the Build-A-Bot platform is built to provide.

The Build-A-Bot platform is currently limited by a relatively small number of 3D robot parts available to choose from inside the application. While there are a number of pre-built robot designs available on the internet that we could use, most of these come as a fully assembled robot. This would significantly limit user creativity and the models we can build. We therefore are creating all the 3D models of robot parts on our own which is a time-consuming process. We are exploring multiple options for how to address this challenge. One possible solution for this issue would be to enable an advanced function where user submitted parts could potentially be used. This solution comes with limitations of its own, as our team would still need to validate the parts submitted before they could be used in our platform.

VI. Future Work

A. Improving Robotic Designs Using Human-Centered AI

Human-Centered Artificial Intelligence is a specialization focused on bridging the gap between humans and machines by developing intelligent systems that can understand how humans perceive and interact with the world around them. As part of this project we will develop a Human-Centered AI approach to processing the design data collected as part of the Build-A-Bot platform. Specifically, we will be turning to Machine Learning (ML) to help us better understand how robot mind perception is causally tied to robot design.

Our data consists of both images of the created robot designs as well as the low-level model information including features such as what model was selected, where it was attached to the design, and what rotation or scale was applied. We will create models based on the pixel information provided by the screenshots of the images as well as models based on the model's composition, and compare the accuracy of these models to see which can be used as a better predictor of RToM for a given design. Our hypothesis during this comparison is that the pixel values can serve as a better predictor of whether a new design conforms to a given RToM target, while the model composition data will serve as a better training dataset for models used to generate new designs. We will also look at grouping models by the relative complexity of the training examples to test the hypothesis that ML models focused on a given complexity will be more accurate than ones trained with a mixture of complexity values.

In order to allow for our Human-Centered AI system to continuously improve itself, all models developed as part of the project will be updated in real-time as new designs are submitted to the system.

B. Validating RToM Perception Using Neuroscience

Independent of the platform design results, we are developing implicit and explicit measures to build a model of robot mind perception. The novel metrics are used to assess and verify the outcomes from the platform robot designs as well as to assess and refine the ML models. To date, robot mind perception research is in its infancy and has no models based on a large dataset and implicit measures to verify their validity. Additionally, no prior research attempted to use implicit measures to link robot design features to robot mind attribution. The interactive testbed we are developing includes explicit measures (e.g. questionnaires), as well as neuroscientific measures (i.e. fNIRS). Functional Near-Infrared Spectroscopy (fNIRS) explores the functional activation of the human cerebral cortex through optical topography. It is non-invasive, silent, low-cost, portable, allows for participant movements, and has good temporal resolution which is highly desired as we research responses to a stimuli. Studies using physiological measures have shown that Functional Near-Infrared Spectroscopy (fNIRS) is suitable to pick up on empathy modulation [23]. fNIRS has also been mentioned to be a physiological method necessary in future HRI studies to determine how social robots should be designed to best perceive user's needs [48]. In order to investigate how fNIRS can help reinforce our findings into RToM produced from the machine learning models training on data provided by the Build-A-Bot platform, we are planning to integrating prior work on fNIRS and HRI to build a new experimental approach in order to establish a proof of concept for measuring RToM.

Our study will focus on investigating an event-related reaction of participants' empathy reactions to human and robot faces displayed in a painful condition (i.e. with a needle penetrating the skin) and in a non-painful condition (i.e. with Q-tip touching the skin). A repeated measure ANOVA will be performed to determine the effect of the touch condition on the fNIRS results, which will let us determine if there is a difference in activation between a painful touch and a pleasant touch. This will help us determine if fNIRS can provide insights into RToM that we can use as a validation method for the findings from our machine learning model, and provide a novel experimental method into RToM that can lead to significant new insights.

VII. Conclusion

We have successfully developed a comprehensive experimental design that has the potential to significantly increase knowledge on how people develop a Robot Theory of Mind (RToM). We created a web-based based platform that will collect a large amount of novel robot designs associated with a mental state. We will be able to determine what mental states are ascribed to robots and how a robot needs to be designed in order to display a certain mental state. In the future, this work will utilize machine learning and neuroscience to significantly contribute to the knowledge in each respective field and give insights on a more comprehensive assessment of interactions with social robots.

References

[1] Glenda Andrews, Graeme S Halford, Katie M Bunch, Darryl Bowden, and Toni Jones. Theory of mind and relational complexity. *Child development*, 74(5):1476–1499, 2003.

[2] Jaime Banks. Theory of mind in social robots: replication of five established human tests. *International Journal of Social Robotics*, 12(2):403–414, 2020.

[3] Simon Baron-Cohen, Alan M Leslie, and Uta Frith. Does the autistic child have a "theory of mind"? *Cognition*, 21(1):37–46, 1985.

[4] Kirsten Bergmann, Friederike Eyssel, and Stefan Kopp. A second chance to make a first impression? how appearance and nonverbal behavior affect perceived warmth and competence of virtual agents over time. In *International conference on intelligent virtual agents*, pages 126–138. Springer, 2012.

[5] Markus Blut, Cheng Wang, Nancy V Wünderlich, and Christian Brock. Understanding anthropomorphism in service provision: a meta-analysis of physical robots, chatbots, and other ai. *Journal of the Academy of Marketing Science*, 49(4):632–658, 2021.

[6] Elizabeth Broadbent, Vinayak Kumar, Xingyan Li, John Sollers 3rd, Rebecca Q Stafford, Bruce A MacDonald, and Daniel M Wegner. Robots with display screens: a robot with a more humanlike face display is perceived to have more mind and a better personality. *PloS one*, 8 (8):e72589, 2013.

[7] Cody Canning and Matthias Scheutz. Functional near-infrared spectroscopy in human-robot interaction. *Journal of Human-Robot Interaction*, 2(3):62–84, 2013.

[8] Boyuan Chen, Carl Vondrick, and Hod Lipson. Visual behavior modelling for robotic theory of mind. *Scientific Reports*, 11(1):1–14, 2021.

[9] Emily S Cross, Katie A Riddoch, Jaydan Pratts, Simon Titone, Bishakha Chaudhury, and Ruud Hortensius. A neurocognitive investigation of the impact of socializing with a robot on empathy for pain. *Philosophical Transactions of the Royal Society B*, 374(1771):20180034, 2019.

[10] Charles R Crowell, Jason C Deska, Michael Villano, Julaine Zenk, and John T Roddy Jr. Anthropomorphism of robots: study of appearance and agency. *JMIR human factors*, 6(2):e12629, 2019.

[11] Carl F DiSalvo, Francine Gemperle, Jodi Forlizzi, and Sara Kiesler. All robots are not created equal: the design and perception of humanoid robot heads. In *Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 321–326, 2002.

[12] Friederike Eyssel, Dieta Kuchenbrandt, Frank Hegel, and Laura De Ruiter. Activating elicited agent knowledge: How robot and user features shape the perception of social robots. In *2012 IEEE RO-MAN: The 21st IEEE international symposium on robot and human interactive communication*, pages 851–857. IEEE, 2012.

[13] Susan T Fiske, Amy JC Cuddy, and Peter Glick. Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences*, 11(2):77–83, 2007.

[14] Marlena R Fraune. Our robots, our team: Robot anthropomorphism moderates group effects in human–robot teams. *Frontiers in Psychology*, 11:1275, 2020.

[15] Jennifer Goetz, Sara Kiesler, and Aaron Powers. Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003.*, pages 55–60. Ieee, 2003.

[16] Heather M Gray, Kurt Gray, and Daniel M Wegner. Dimensions of mind perception. *science*, 315(5812):619–619, 2007.

[17] Kurt Gray and Daniel M Wegner. Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, 125(1):125–130, 2012.

[18] Kurt Gray, Liane Young, and Adam Waytz. Mind perception is the essence of morality. *Psychological inquiry*, 23(2):101–124, 2012.

[19] Kerstin S Haring, Katsumi Watanabe, Mari Velonaki, Chad C Tossell, and Victor Finomore. FFAB-The form function attribution bias in human-robot interaction. *IEEE Transactions on Cognitive and Developmental Systems*, 10(4):843–851, 2018.

[20] Nick Haslam. Dehumanization: An integrative review. *Personality and social psychology review*, 10(3):252–264, 2006.

[21] Frank Hegel, Claudia Muhl, Britta Wrede, Martina Hielscher-Fastabend, and Gerhard Sagerer. Understanding social robots. In *2009 Second International Conferences on Advances in Computer-Human Interactions*, pages 169–174. IEEE, 2009.

[22] Cecilia M Heyes. Theory of mind in nonhuman primates. *Behavioral and brain sciences*, 21(1):101–114, 1998.

[23] Toshiyuki Himichi and Michio Nomura. Modulation of empathy in the left ventrolateral prefrontal cortex facilitates altruistic behavior: An fnirs study. *Journal of Integrative Neuroscience*, 14(02):207–222, 2015.

[24] Daniel Kahneman. *Thinking, fast and slow*. Macmillan, 2011.

[25] Min-Sun Kim and Eun-Joo Kim. Humanoid robots as "the cultural other": are we able to love our creations? *AI & society*, 28(3):309–318, 2013.

[26] Wonjoon Kim, Nayoung Kim, Joseph B Lyons, and Chang S Nam. Factors affecting trust in high-vulnerability human-robot interaction contexts: A structural equation modelling approach. *Applied ergonomics*, 85:103056, 2020.

[27] Michael Lewis, Katia Sycara, and Phillip Walker. The role of trust in human-robot interaction. In *Foundations of trusted autonomy*, pages 135–159. Springer, Cham, 2018.

[28] Sunny Xun Liu, Qi Shen, and Jeff Hancock. Can a social robot be too warm or too competent? older chinese adults' perceptions of social robots and vulnerabilities. *Computers in Human Behavior*, 125:106942, 2021.

[29] Barbara CN Müller, Xin Gao, Sari RR Nijssen, and Tom GE Damen. I, robot: How human appearance and mind attribution relate to the perceived danger of robots. *International Journal of Social Robotics*, 13(4):691–701, 2021.

[30] Manisha Natarajan and Matthew Gombolay. Effects of anthropomorphism and accountability on trust in human robot interaction. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 33–42, 2020.

[31] Elizabeth Phillips, Scott Ososky, Janna Grove, and Florian Jentsch. From tools to teammates: Toward the development of appropriate mental models for intelligent robots. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 55, pages 1491–1495. SAGE Publications Sage CA: Los Angeles, CA, 2011.

[32] Elizabeth Phillips, Xuan Zhao, Daniel Ullman, and Bertram F Malle. What is human-like? decomposing robots' human-like appearance using the anthropomorphic robot (abot) database. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*, pages 105–113, 2018.

[33] Aaron Powers and Sara Kiesler. The advisor robot: tracing people's mental model from a robot's physical attributes. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 218–225, 2006.

[34] 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 *Annual Conference Towards Autonomous Robotic Systems*, pages 349–361. Springer, 2019.

[35] 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. IEEE, 2020.

[36] Alessandra Rossi, Patrick Holthaus, Giulia Perugia, Sílvia Moros, and Marcus Scheunemann. Trust, acceptance and social cues in human–robot interaction (scrita), 2021.

[37] Silvia Rossi, Daniela Conti, Federica Garramone, Gabriella Santangelo, Mariacarla Staffa, Simone Varrasi, and Alessandro Di Nuovo. The role of personality factors and empathy in the acceptance and performance of a social robot for psychometric evaluations. *Robotics*, 9 (2):39, 2020.

[38] Brian Scassellati. Theory of mind for a humanoid robot. *Autonomous Robots*, 12(1):13–24, 2002.

[39] Sara M Schaafsma, Donald W Pfaff, Robert P Spunt, and Ralph Adolphs. Deconstructing and reconstructing theory of mind. *Trends in cognitive sciences*, 19(2):65–72, 2015.

[40] Kristin Schaefer. The perception and measurement of human-robot trust. 2013.

[41] Kristin E Schaefer, Tracy L Sanders, Ryan E Yordon, Deborah R Billings, and Peter A Hancock. Classification of robot form: Factors predicting perceived trustworthiness. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 56, pages 1548–1552. SAGE Publications Sage CA: Los Angeles, CA, 2012.

[42] Stela H Seo, Denise Geiskkovitch, Masayuki Nakane, Corey King, and James E Young. 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)*, pages 125–132. IEEE, 2015.

[43] Nicolas Spatola and Olga A Wudarczyk. Ascribing emotions to robots: Explicit and implicit attribution of emotions and perceived robot anthropomorphism. *Computers in Human Behavior*, 124:106934, 2021.

[44] Yutaka Suzuki, Lisa Galli, Ayaka Ikeda, Shoji Itakura, and Michiteru Kitazaki. Measuring empathy for human and robot hand pain using electroencephalography. *Scientific reports*, 5(1):1–9, 2015.

[45] Haodan Tan, Dakuo Wang, and Selma Sabanovic. Projecting life onto robots: The effects of cultural factors and design type on multi-level evaluations of robot anthropomorphism. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 129–136. IEEE, 2018.

[46] Unity Technologies. Unity Real-Time Development Platform | 3D, 2D VR & AR Engine. URL https://unity.com/.

[47] Adam Waytz and Michael I Norton. Botsourcing and outsourcing: Robot, british, chinese, and german workers are for thinking – not feeling – jobs. *Emotion*, 14(2):434, 2014.

[48] Eva Wiese, Giorgio Metta, and Agnieszka Wykowska. Robots as intentional agents: using neuroscientific methods to make robots appear more social. *Frontiers in psychology*, 8:1663, 2017.

[49] Liane Young and Adam Waytz. Mind attribution is for morality. *Understanding other minds: Perspectives from developmental social neuroscience*, pages 93–103, 2013.

[50] Jakub Złotowski, Hidenobu Sumioka, Shuichi Nishio, Dylan F Glas, Christoph Bartneck, and Hiroshi Ishiguro. Appearance of a robot affects the impact of its behaviour on perceived trustworthiness and empathy. *Paladyn, Journal of Behavioral Robotics*, 7(1), 2016.