A Hierarchical Model for Human-Robot Interaction

Barfield, Jessica K. School of Information Sciences, University of Tennessee-Knoxville | jbarfiel@vols.utk.edu

ABSTRACT

This paper proposes a hierarchical model for human interaction with social robots which consists of three tiers, each tier based on an existing theory of human-robot interaction. The three theories which comprise the model are robot anthropomorphism which, among others, is based on the physical appearance of the robot; Social Identity Theory, which discusses how a robot may be classified into social categories; and at the lowest tier of the model, the Computers as Social Actors Theory, in which social rules mediate human-robot interaction. The top level of the model is based on the initial impression that individuals form of robots which is determined primarily by the robot's physical appearance. At the mid-tier of the model, Social Identity Theory is used to describe how people place robots into social categories, which among others can be based on the perceived race, ethnicity, and gender of the robot. An important consequence of categorizing robots by a social class is that the category may determine whether the robot is judged to be an in-group or out-group member which can lead to different behavioral interactions between humans and robots as described in the third tier of the model. According to the hierarchical model, communication with a robot occurs at the lowest tier of the model and is guided by the process of robot anthropomorphism, Social Identity Theory, and by predictions from the Computers as Social Actor paradigm.

KEYWORDS

Social Identity Theory; human-robot interaction; robot anthropomorphism; CASA, hierarchical model

INTRODUCTION

This paper proposes a hierarchical model for human-robot interaction (HRI) which consists of three-tiers that together describe the process of communication between humans and robots as experienced in social contexts. The hierarchical model is based on existing stand-alone theories which describe different aspects of human interaction with robots. When considered together, the three theories provide a comprehensive framework in which to view HRI from the initial reaction of the robot to the level of individual communication. Further, the hierarchical model implies an order to the process of human-robot interaction and how communication with a robot may be mediated by robot anthropomorphism and by the social categorization of the robot. At the highest level of the model, an initial impression of a robot is formed based on the process of "robot anthropomorphism" which occurs when people initially ascribe human characteristics to the robot based on the robot's physical appearance (Spatola et al., 2022). Following the initial impression of the robot, in the middle stage of the hierarchical model, people access knowledge from long-term memory which they use to place robots into different social categories, a process which is described by Social Identity Theory (SIT) (see Meltzoff and Moore, 2002). Finally, at the lowest tier of the model, based on input provided by the social categorization stage of the model, the Computers as Social Actors paradigm (CASA) (Nass, Steuer, and Tauber, 1994) is used to explain how people engage in specific types of behavior in social interactions with a robot. Thus, CASA focuses on the rules people use to interact with robots once they are anthropomorphized and categorized into social groups. At this level of abstraction in HRI, human performance for applied tasks such as information search and retrieval is dependent on how robots are processed in the preceding tiers of the model.

The hierarchical model for HRI proposes that communication with a robot follows an ordered process from an initial impression of the robot to a lower level of abstraction where individuals communicate with the robot in social contexts. Such communication is dependent on the higher-level components of the model thus indicating that individual communications with social robots are part of a process that starts before verbal exchanges occur. Thus, as an implication, designers of robots may influence communication with a robot by its physical appearance and its similarity to a user which, among others, determines the social category the robot is placed. Further, while the model is shown as linear in structure in that the output of one tier of the model feeds into another; still, there can be overlap between the different tiers of the model. Thus, the model should be considered a broad conceptual framework in which to guide research on HRI. Summarizing, the objective of the paper is to describe a model which accounts for the initial impression formed of a robot, how a robot is categorized into a social class, and how these aspects of HRI influence the communicative behavior between human and robot. From this approach it is expected that a hierarchical model of HRI which is inclusive as compared to previous singular models of HRI can be used to evaluate and design future studies and to help formulate research questions to explain how users interact with social robots. Figure 1 shows the model which is explained more fully in the following sections of the paper.

HIERARCHICIAL HUMAN-ROBOT INTERACTION MODEL

TIER 1: Robot Anthropomorphism

As AI-enabled and humanoid robots enter society, they are beginning to interact with people in different social contexts in which they display a range of social skills. As shown by the process of anthropomorphism, past research

Mid-Year Conference of the Association for Information Science & Technology | April 11-13, 2023, Virtual Event.

on HRI has found that people assign human characteristics to robots based primarily on the robot's physical appearance (Blut et al., 2021). And if the robot's physical appearance resembles a human but not quite reaching humanness in appearance, robots that imperfectly resemble human beings will provoke uncanny feelings of uneasiness or even of revulsion in observers (Kim, de Visser, and Phillips, 2022). That people anthropomorphize robots has been shown in numerous studies and suggests that if a person interacts with a robot that has attributes which differ from those of the observer, the person may react in a stereotypical and biased manner toward the robot (Bartneck et. al., 2009; Epley, Waytz, and Cacioppo, 2007; Eyssel and Kuchenbrandt, 2012; Deligianis, et al. 2017; Kamide, Eyssel, and Arai, 2013). In the hierarchical model presented here, I propose that as an initial reaction to a robot experienced in a social context, the robot is anthropomorphized based on its physical appearance, which has consequences for the next two tiers of the model. Of interest for HRI, robot anthropomorphism can be manipulated for various reasons. For example, van Pinxteren et al. (2019) studying service robots observed that firms often include human-like features in service robots to increase trust (which would be displayed in tier 3 of the model). In the hierarchical model shown below, the process of robot anthropomorphism leads to the second tier of the model which is described next. The triggering mechanism, or transition, from tier 1 to tier 2 is the individual's initial reaction to the physical structure of the robot.

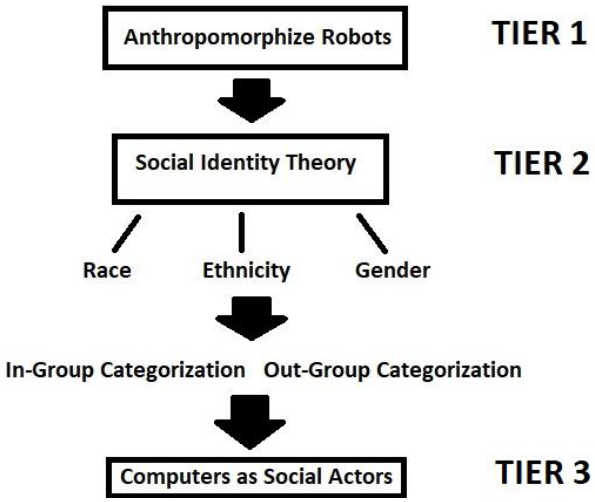

Figure 1. Conceptual framework for human-robot interaction.

TIER 2: Social Identity Theory

The hierarchical HRI model discussed in this paper states that after a robot is anthropomorphized based primarily on its physical appearance, an individual will further process the robot's features and behavior to determine whether the robot matches or not the user's social characteristics. That is, in the second stage of the model, a robot is categorized into a social class which has implications for the classifier and for HRI. At this level of HRI, Social Identity Theory (SIT) proposes that individuals gain a sense of personal worth from the groups with which they identify and that they prefer to interact with group members that have similar characteristics to them (Tajfel, 2010). Social Identity Theory also proposes that people with dissimilar characteristics may be viewed as an out-group member, and those with similar characteristics viewed as an in-group member (Eyssel and Kuchenbrandt, 2012).

Group membership may depend on a number of factors. In a study evaluating the social categorization of robots, Eyssel and Hegel (2012) found that there is a tendency to place robots into racial, gender, and ethnic categories (see also Eyssel and Loughnan, 2013; Li et al. 2022; Nass and Brave, 2005). For example, Eyssel and Kuchenbrandt (2012) investigating robot ethnicity used German citizens to evaluate robots that were given either a Turkish or German identity. Their objective was to determine whether the robots that differed only by ethnic name and national origin was sufficient to signal a German or Turkish identity and subsequent differential treatment. The results of their study indicated that a robot introduced to the German citizens as a Turkish product compared to the same robot introduced to subjects as a German product, received a significantly more negative evaluation among German citizens. In fact, a number of social psychology studies have concluded that in-group or out-group bias, and thus preferential or non-preferential treatment, can be triggered by markers of similarity between group members (Turner, 1978). Illustrating this point, Bartneck et al., (2018) showed that robots designed with different surface colors, were "racialized" by observers in that robots perceived as a member of the observer's "out-group" were judged less favorably than robots perceived as a member of the observer's "in-group" (see generally Barfield, 2021; Eyssel and Loughnan, 2013). Further, whether a robot is viewed as an in-group or out-group member may determine

how it is accepted by individuals and its subsequent treatment as discussed in tier 3 of the model. To proceed to tier 3 of the model, I propose that the user has placed the robot into a social class based on similarities to the user.

TIER 3: Computers as Social Actors Paradigm

The third tier of the model describing HRI is represented by the CASA paradigm. The CASA paradigm (Nass and Moon, 2000; Nass et al., 1994) was derived from the media equation proposed by Reeves and Nass (1996), and suggests that people treat robots like people, and for purposes of the model discussed here, apply scripts used for interacting with humans to interactions with robots that are experienced in social contexts. So, in tier 2 of the hierarchical model I proposed that people place robots into social categories, and in tier 3 the social category that the robot is placed in influences individual behaviors towards the robot. For example, in tier 2, the social categorization process may group robots into in-group or out-group members and research has shown that in-group members are treated more favorably (a process which occurs in tier 3). On this point, from the CASA paradigm it has been shown that humans interact with robots in a similar manner as they do other people (Nass, Steuer, and Tauber, 1994). As an explanation of this phenomenon for HRI, Eyssel and Kuchenbrandt (2012) commented that people draw on their own self-knowledge, or their knowledge about other people, when judging non-human entities. Of course, CASA doesn't predict that interactions between humans and robots will be positive, just that depending on the social circumstances the interaction will be similar to how humans interact with each other. On this point, the results of several studies have shown that as people interact with robots, they may express biases toward the robot not unlike the biases that people of color, or of certain ethnic groups, or of a particular gender currently receive in society (Barfield, 2021; Edwards et al. 2019; Eyssel and Loughnan, 2013; Keijsers and Bartneck, 2018; Louine, et. al., 2018).

Much of what is considered social behavior occurs in tier 3 of the model and therefore has direct consequences for HRI. For example, with robots becoming more humanoid in appearance and equipped with AI, there has been a growing effort to investigate whether people discriminate against robots and respond to them based on their categorization by gender, ethnicity, or race (Barfield, 2021). For example, Eyssel and Hegel (2012) investigating the effect of facial cues on the perception of robot gender asked whether a robot designed as gendered female would be stereotyped as female, and whether a robot designed as gendered male would be stereotyped as male. The findings indicated that the same gender stereotypes which bias social perceptions of humans, are also applied to robots. For example, "male appearing" robots were ascribed more agency-related traits, and "female appearing" robots were ascribed more communal traits (Eyssel and Hegal, 2012). More recently, Otterbacher and Talias (2017) found that people responded to robots using stereotypical responses thought to be representative of a particular gender. Specifically, participant's evaluations of the female gendered robots were categorized as being emotionally warm and the male gendered robots as being more agentic (Otterbacher and Talias, 2017).

Given the discussion of the literature and description of the hierarchical model, how could such a model be tested? Clearly, the individual models (anthropomorphism, SIT, CASA) have been the subject of comprehensive research so, a question to pose is what triggers movement from one tier of the model to another? On this question, as a research approach it would be useful to propose "triggers" for such movements and then to evaluate whether they led to the next tier. Of course, with a three-tier model, that process can occur between tiers 1 and 2, and between tiers 2 and 3. For example, tier 2 predicts that individuals will categorize robots as either an in-group or out-group member, and when the social categorization process is completed should lead to communication patterns in tier 3 consistent with the social categorization.

CONCLUSIONS: TOWARDS AN INTEGRATED MODEL OF HRI

Human-robot interaction is a complex process in which humans continue to evaluate robots throughout the entire process of interacting with them. Therefore, a model which considers the various stages of HRI as users gain familiarity with the robot should have considerable value for the HRI community and in providing a conceptual framework in which to evaluate HRI studies and to generate research questions. The hierarchical model I propose for HRI which as shown in Figure 1 consists of three tiers in HRI accounting for the initial exposure to robots to the level of individual exchanges of information between human and robot.

Further, the model combines different theories of HRI that individually consider important aspects of interacting with robots in social contexts. However, for further development of the model, since research on HRI has shown that results are often task specific, it is necessary to test how well the model generalizes across different tasks and domains. At any rate, a motivation for the hierarchical model is the need for a more comprehensive approach that explains various aspects of HRI which covers initial contact with a robot, to processes which occur during a more extended time period interacting with robots in which the user becomes more familiar with the robot's physical form and behaviors. For example, anthropomorphism occurs when individuals assign human-like affordances to a robot based on the physical appearance of the robot, and Social Identity Theory proposes that individuals place robots into social categories dependent among others, on how the robot is anthropomorphized. Further, CASA indicates that

humans may interact with robots in a comparable manner as they do other people. While the above model is presented as sequential, after the initial impression of a robot, I do not think that each stage must be completed before the next stage begins (thus there can be overlap in the stages), but I do think the events which happen in each stage of the model follow a logical progression for HRI which I glean from my own research, from past studies on robot anthropomorphism, from research on the social categorization of robots (SIT), and from the literature on how people interact with robots in social contexts (CASA).

To reiterate, each component of the hierarchical model helps to explain HRI at a particular stage in the interaction between humans and robots, and when combined as a three-tier process results in a more comprehensive description of HRI than expressed in previous models and theories. For sake of simplicity, I now refer to the model as the A(anthropomorphism), S(social classification), and C(communication), or ASC model. The value of the ASC model is that it is comprehensive enough in coverage to account for the initial reaction to a robot, its categorization into a social class (matching or not that of the user), and how the prior categorization of a robot influences the process of human-robot interaction. In summary, the ASC model contributes to theory in HRI, and also has practical significance in that the three tiers described in the model have implications for HRI at the level of the process of communication between human and robot. Finally, future studies by the author will be done to test the overall predictive nature of the model, and studies will be run to determine how social classification of a robot (tier 2) influences the particular exchange of information which I propose occurs in tier 3 of the ASC model.

ACKNOWLEDGEMENT

The author acknowledges support from the University of Tennessee at Knoxville during the writing of this paper.

REFERENCES

Barfield, J. K., Discrimination and Stereotypical Responses to Robots as a Function of Robot Colorization, Adjunct Proceedings of the 29th ACM Conference on Modeling and Personalization, 109-114, 2021.

Bartneck, C., Croft, E., Kulic, D., & Zoghbi, S. (2009). Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots, *International Journal of Social Robotics*, Vol. 1, 71–81.

Bartneck, C., Yogeeswaran, K., Ser. Q. M., Woodward, G., Sparrow, R., Wang, S., & Eyssel, F. (2018). Robots and Racism, *HRI '18: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 196–204.

Blut, M., Wang, C., Wunderlich, N. V., & Brock, C. (2021). Understanding Anthropomorphism in Service Provision: a Meta-Analysis of Physical Robots, Chatbots, and Other AI, *Journal of the Academy of Marketing Science*, Vol. 49 (4), 632-658.

Deligianis, C., Stanton, C., McGarty, C., & Stevens, C. J. (2017). The Impact of Intergroup Bias on Trust and Approach Behaviour Towards a Humanoid Robot, *Journal of Human-Robot Interaction*, Vol. 6 (3), 4-20.

Edwards, C., Edwards, A., Stoll. B., Lin., X., & Massey, N. (2019). Evaluations of an Artificial Intelligence Instructor's Voice: Social Identity Theory in Human-Robot Interactions, *Computers in Human Behavio*r, Vol. 90, 357-362.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On Seeing Human: A Three-Factor Theory of Anthropomorphism, *Psychological Review*, Vol. 114, 864-886.

Eyssel, F., & Hegel, F. (2012). (S)he's Got the Look: Gender Stereotyping of Robots, *Journal of Applied Social Psychology*, Vol. 42(9), 2213-2230.

Eyssel, F., & Kuchenbrandt, D. (2012), Social Categorization of Social Robots: Anthropomorphism as a Function of Robot Group Membership, *British Journal of Social Psychology*, Vol. 51, 724-731.

Eyssel, F., & Loughnan, S. (2013). It Don't Matter if You're Black or White? Effects of Robot Appearance and User Prejudice on Evaluations of a Newly Developed Robot Companion, 422-433, In: Herrmann G., Pearson M. J., Lenz A., Bremner, P., Spiers A., & Leonards U. (eds) *Social Robotics, ICSR 2013, Lecture Notes in Computer Science*, Vol. 8239. Springer, Cham.

Kamide H., Eyssel F., & Arai T. (2013). Psychological Anthropomorphism of Robots. In: Herrmann G., Pearson M. J., Lenz A., Bremner P., Spiers A., & Leonards U. (eds), *Social Robotics, ICSR 201, Lecture Notes in Computer Science*, Vol. 8239. Springer, Cham.

Keijsers, M., & Bartneck, C. (2018). Mindless Robots Get Bullied, *Proceedings of the ACM/IEEE International Conference on Human Robot Interaction*, Chicago, 205-214.

Kim, B; de Visser, E., & Phillips, E. (2022). Two Uncanny Valleys: Re-Evaluating the Uncanny Valley Across the Full Spectrum of Real-World Human-Like Robots, *Computers in Human Behavior,* Vol. 135.

Li, L., Li, Y., Song, S., Zhaomin, S., & Wang, C. L. (2022). How Human-like Behavior of Service Robots Affects Social Distance: A Mediation Model and Cross-Cultural Comparison, *Behavioral Sciences*, Vol. 12 (7), 205.

Louine, J., May, D. C., Carruth, D. W., Bethel, C. L., Strawderman, L., & Usher, J. M. (2018). Are Black Robots Like Black People? Examining How Negative Stigmas about Race are Applied to Colored Robots, *Sociological Inquiry*, Vol. 88(4), 626-648.

Meltzoff, A. N., & Moore, M. K. (2002). Imitation, Memory, and the Representation of Persons, *Infant Behavior & Development*, Vol. 25 (1), 39-61.

Nass, C., & Brave, S. (2005). *Wired for Speech*, MIT Press.

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers, *Journal of Social Issues*, Vol. 56, 81–103.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are Social Actors, *Proceedings of SIGCHI '94 Human Factors in Computing Systems*, 72–78.

Otterbacher, J., & Talias, M. (2017). S/he's too Warm/Agentic! The Influence of Gender on Uncanny Reactions to Robots, *HRI'17 Conference*, 214-223.

Reeves, B., & Nass, C. (1996). The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge University Press.

Spatola, N., Marchesi, S., & Wykowska, A. (2022). The Phenotypes of Anthropomorphism and the Link to Personality Traits, *International Journal of Social Robotics*.

Tajfel, H. (2010). *Social Identity and Intergroup Relations (European Studies in Social Psychology, Series Number 7)*, Cambridge University Press, Reissue edition.

Turner, J. C. (1978). Social Comparison, Similarity, and Ingroup Favoritism, In H. Tajfel (ed.), *Differentiation Between Social Groups: Studies in the Social Psychology of Intergroup Relations*, Academic Press.

van Pinxteren, M. M. E., Wetzels, R. W. H., Ruger, J. (2019). Pluymaekers, M., and Wetzels, M., Trust in Humanoid Robots: Implications for Services Marketing, *Journal of Services of Marketing*, Vol. 33 (4), 507-518.