Towards a Common Understanding and Vision for Theory-Grounded Human-Robot Interaction (THEORIA)

Glenda Hannibal
TU Wien
Vienna, Austria
glenda.hannibal@tuwien.ac.at

Nicholas Rabb
Tufts University
Medford, Massachusetts, United States
nicholas.rabb@tufts.edu

Theresa Law
Tufts University
Medford, Massachusetts, United States
theresa.law@tufts.edu

Patrícia Alves-Oliveira
University of Washington
Seattle, Washington, United States
patri@cs.washington.edu

Abstract—**While the accumulation of practical knowledge provided researchers with much insight into successful human-robot interaction (HRI), a broader discussion about the role of *theoretical knowledge* is still lacking. It is unfortunate because it is also important to explicitly consider theory and theorizing in HRI as crucial contributions when aspiring to develop this field of research into a mature science. With our proposed interactive half-day workshop, we aim to provide a vibrant setting for the participants to discuss the "what, why, and how" of theoretical knowledge in HRI, as they share and learn from each other's experiences and competence. In the long-term perspective, the outcome of this workshop will lay the foundation for a supportive *research community* that encourages researchers to reflect and collaborate further on theory-grounded HRI work.**

Index Terms—**human-robot interaction, knowledge, theory, philosophy of science, interactive workshop, community building**

I. Theme

Aristotle famously opened his work on Metaphysics with the statement that "all men by nature desire to know" [1]. From his careful reflections on the different ways of knowing, the distinction between *theoretical knowledge* and *practical knowledge* has been a central theme in epistemology and the philosophy of science ever since ancient times. Awareness of how theoretical and practical knowledge play different, but interrelated, roles in the growth of knowledge is important to any field of research that aims to become a mature science. While practical knowledge enables understanding through hands-on and personal experience (i.e., acts of *doing*), theoretical knowledge provides a fundamental understanding of the principles behind a subject matter in relation to a greater whole (i.e., acts of *reasoning*). Both are essential for the growth of HRI as a field of research.

Practical knowledge currently dominates human-robot interaction (HRI) study, as countless studies on what makes communication or interactions between humans and robots

successful or not, and careful hypothesis testing of various established social theories comprise the impressive body of literature on HRI as it stands today. These insights have led to the formation of an independent field of research in robotics with well-established annual conferences (e.g., IEEE RO-MAN, ACM/IEEE HRI, Springer ICSR), dedicated journals (e.g., Transactions in HRI, International Journal of Social Robotics, Interaction Studies), and several handbooks (e.g., [2], [6]). While the practical application of social theories in HRI has gained recognition for advancing the development of socially capable robots and supporting a better understanding of their acceptance and use in society more broadly, it is becoming more apparent that discussions about the importance of theoretical knowledge in HRI has been neglected. To this day, there have been few attempts in the literature to explicitly reflect on the value of theoretical knowledge in HRI as a way to push this field of research into a mature science [3], [5], [7]–[9].

Overall, the main concerns when discussing the importance of theoretical knowledge in the field of HRI have focused on clarifying central concepts, and reflecting diligently on methodological issues required to improve an understanding of the complex exchange between humans and robots. Some researchers have sought to clarify paradigms used in HRI by grounding them in the literature of their respective home fields [5], [8]. Others have argued that robots present such novelty that theory from human-based study should not be naively ported over to HRI [3], [7]. Even methodology is analyzed, as some claim that watered-down laboratory studies cannot capture the complexity that is necessary for solid HRI theory [9].

Without a synthesis of practical and theoretical knowledge, HRI stands to potentially develop into a field with imprecise concepts. Relying only on results from empirical studies are not sufficient to understand the implications of robots in

our lives and societies, including theorizing and getting a deeper understanding of the phenomena. A solely results-driven field, using measures from mature sciences, risks being one that does not account for what is actually being done in the process. Implications of research may not be understood without a rigorous grasp of underlying concepts. There are both epistemological and real-world dangers to such a field.

II. Aim

The HRI'22 conference track on "Theory and Methods" describes contributions to the development of theoretical and practical knowledge as, "new ways of studying HRI, elucidating or connecting fundamental HRI principles beyond individual interfaces or projects, new theoretical concepts in HRI, literature reviews, work that focuses on reproducing, replicating, or recreating prior HRI work (or fails to), etc."[1]. While the intention behind such track is to value both types of knowledge equally, there seems to be a tendency to downplay contributions that are solely theory-based. We posit that this imbalance arises because there is not currently a common understanding of what theory and theorizing means in HRI. In our view, insisting on having more explicit discussions in the HRI community about how to create a theory, improve an existing one, or how to evaluate the quality of a theory is not only needed to better support study designs, but is also necessary for better facilitation of interdisciplinary collaboration.

To reach the goal of creating a recursive feedback loop between theoretical and practical knowledge, this interactive workshop sets out to explore theory-grounded HRI by jointly discussing what it could and should mean. The claim is not that theory or theorizing takes priority or always has to be the starting point for HRI research – only that the contribution of theoretical knowledge deserves more attention because it "defines the variables, specifies the domain, builds internally consistent relationships, and makes specific predictions" [10, 361]. Through participating in this workshop, both young and senior researchers will have the opportunity to immerse themselves in questions about what theory-grounded HRI research is. They will do this by engaging in critical dialogue about theory and theorizing in HRI, expanding upon and discussing their understanding and application of theoretical knowledge in their own work, and making connections with others.

The aim of this workshop is two-fold. Firstly, we *highlight the importance of theory and theorizing for the further development of HRI* as a mature field of research. Secondly, we believe that by providing a venue focused on theory-grounded HRI, we *create and foster a community* around this perspective that can grow into a strong movement within the field.

III. Topics

Researchers who are interested in theory-grounded work in HRI can join us in discussion about topics including, but not limited to:

---

[1]Retrieved from the ACM/IEEE HRI Conference official website: https://humanrobotinteraction.org/2022/full-papers/

1) The "What", "Why" , and "How" of theory and theorizing. Specifically questions about (i) deriving theory from data, creating theory-driven studies, (ii) understanding the recursive relationship between theory and empirical work, (iii) understanding and application of theories from different disciplines
2) Specific application of theory to HRI. Specifically questions about (i) the state of the art in HRI related to theory/theorizing, (ii) identifying different candidates for "theories" resulting from HRI work, (iii) the value of, and justification for, theory and theorizing in HRI.

IV. Audience

We welcome anyone who is interested in discussing the above topics, and it is not a requirement to have any specialized knowledge or skills (e.g. training in programming, statistics, or logic) to participate in the workshop. However, we do expect all participants to have a general overview of the various concepts and methods used in HRI to get the most out of their participation. Our aim is to attract HRI researchers with various disciplinary backgrounds who are curious and eager to engage in discussions about theory and theorizing in HRI. Specifically, we call for researchers who are interested in: (a) developing a community of theory-driven ideas, (b) asking questions about the field: *why are we doing what are we doing, and how do we want to do it?*, (c) testing/applying a theory in HRI and learning how results "feed back" into the theory, (d) speaking about what may be good practices to bring to HRI based on their own experience with other fields that heavily rely on theory.

Researchers interested in participating will have to submit either a short personal statement up to 500 words that present their background and motivation, or a 2-4 page position paper that address central questions about the role of theoretical knowledge for HRI (using the regular conference template and guidelines). Both submission types will undergo a light review by the workshop organizing team. We expect to welcome 20-25 participants to the workshop, which will allow for an interactive, collaborative format.

V. Recruitment

Social media posts (e.g., Twitter, Facebook, LinkedIn), targeted mailing lists (e.g., roboticsworldwide, hri-announcement, philos-l, bcs-hci, eusset), and our own research networks will be used for advertising and recruiting participants for the workshop. We will make these calls for participation not only among HRI researchers, who might already be interested in also attending the conference, but also similar HRI communities (e.g., The Research Network for Transdisciplinary Studies in Social Robotics (TRANSOR), Society for Philosophy of Technology (SPT), The Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB)).

VI. Format

The organizing team aim to *break boundaries* by proposing an unconventional half-day interactive workshop format that

inspires knowledge exchange and discussion about theory and theorizing in HRI. Highly interactive in nature, our choice of workshop format is intended to facilitate group discussion and reflection first and foremost. We believe that such a format will help to begin to build the community of theory-grounded researchers that we want to foster. We are organizing these interactions so that they can be done either virtually or in-person with the help of online facilitation tools. We will structure our workshop as follows:

(i) Welcome: We will start with an introduction section where the organizing team will frame the motivation and purpose behind this workshop with a short presentation. We will also use this time to introduce participants to each other, and begin to build a backdrop conducive for collaborative discussion.

(ii) The "What": To ease into investigating theory with participants, we will center the question: "What is theory and theorizing in your view?" This question will lend insights into HRI researchers are currently considering the role and importance of theoretical knowledge in the field. Building on participants' responses, facilitated through a series of collaborative activities, we will come to a working framework that will help us understand what we mean by theory and theorizing in HRI in the first place.

(iii) The "Why": Unpacking and addressing more general questions of "why" theory and theorizing is necessary for any scientific field. Again, through facilitated conversations and collaborative activities, participants will use their existing expertise regarding why theory is important in other scientific fields to arrive at a working framework that will highlight why theory is important to HRI.

(iv) The "How": Finally, the organizing team will guide participants through using the previously generated frameworks (regarding "what" theory is, and "why" it is important) to comment on "how" theory and theorizing could be applied to further research in specific HRI topics. For the topic of trust in HRI for example, the group could try to analyze what the theories of trust in HRI are. Perhaps the group cannot pin down a definite theory of trust in HRI based on the framework generated during the "what" portion, though they know that having a testable theory is important from the "why" portion. This motivates part of the workshop to naturally conclude with suggestions for how the field of HRI should be envisioned to support a direction towards being more theory-grounded.

VII. TEAM

Glenda Hannibal. Glenda holds a BA and MA in Philosophy from Aarhus University. She worked in the Department of Sociology at the University of Vienna before joining TU Wien in 2018 as a PhD student in the Trust Robots Doctoral College and as a member of the Human-Computer Interaction group. Glenda also works as an expert in the HUMAINT project at the EU Commission. Her research focuses on bringing perspectives from Philosophy together with Social Robotics and HRI mainly in the areas of metaphysical, epistemology,

and philosophy of science. In 2018 Glenda organized a workshop on interdisciplinary research for social robotics at the international conference Robophilosophy, which also resulted in the publication of a special issue [4].

Nicholas Rabb. Nicholas is a PhD candidate studying computer science and cognitive science and a D3M Fellow at Tufts University. He holds a BS in computer science from the Rochester Institute of Technology. His research primarily focuses on using computational methods to study the spread of disinformation and changes in public opinion, and secondarily on the social impacts and critical analysis of digital technologies. Outside of academic work, Nicholas is an involved community organizer, and has held dozens of workshops and talks centered around political education and critical theory.

Theresa Law. Theresa received her BA in cognitive science from Vassar College in 2018. She is currently a PhD candidate studying computer science and cognitive science at Tufts University. Her primary research focus is on social trust in HRI and how that is affected by people's mental models of robots.

Patrícia Alves-Oliveira. Patrícia is a Postdoctoral Research Associate at the University of Washington. Previously she received her Ph.D. from ISCTE-IUL and Cornell University on the topic Creativity and Robotics. Patrícia is broadly interested in studying how robots can be used to improve our lives. She was involved in the organization committees of the HRI Conference 2020, HRI Pioneers 2017, RSS Pioneers 2019, AI for HRI Symposium, among others. She has published in conferences such as HRI, IDC, ICSR, RO-MAN, RSS, and IROS. Patrícia received the Best Paper Award in the HRI Conference in 2020 and 2016.

VIII. DOCUMENTATION

To ensure that the outcome of the workshop is preserved and accessible to both the participants and the HRI community, we will continue the initiative after the conference by taking on the following tasks: (1) Updating the workshop website[2] with a summary of the discussions held. We also intend to keep the website as a venue for establishing collaborative projects among the participants who wish to use this opportunity to work more on theory-grounded HRI in the future. Our long-term plan is to *establish this website as a go-to, online portal* to be informed about the latest publications, symposia, conferences, and workshops on research in HRI that centers theoretical perspectives and theory-building. (2) Using the knowledge gathered from the workshop to *craft a proposal for an open call special issue* with a relevant high-profile journal (e.g., International Journal of Social Robotics, Interaction Studies, Transaction in HRI, AI & Society, Minds and Machines), which we will also encourage all participants to consider contributing to.

[2]Workshop website: https://theoriahri.wixsite.com/theoria

REFERENCES

[1] Aristotle. (1925). *Metaphysics: Book I.* (Translated by W. D. Ross). The Internet Classics Archive: http://classics.mit.edu/Aristotle/metaphysics.1.i.html (original work published 350 B.C.E).

[2] Bartneck, C., Belpaeme, T., Eyssel, F., Kanda, T., Keijsers, M., & Šabanović, S. (2020). *Human-Robot Interaction: An Introduction*. Cambridge, UK: Cambridge University Press.

[3] Fox, J., & Gambino, A. (2021). Relationship Development with Humanoid Social Robots: Applying Interpersonal Theories to Human–Robot Interaction. *Cyberpsychology, Behavior, and Social Networking*, 24(5), 294-299.

[4] Hannibal, G. & Weiss, A. (Eds.) (2020).Envisioning Social Robotics: Current Challenges and New Interdisciplinary Methodologies. *Interaction Studies*, 21(1).

[5] Hannibal, G. Weiss, A. & Charisi, V. (2021). "The robot may not notice my discomfort" – Examining the Experience of Vulnerability for Trust in Human-Robot Interaction. Proceedings of the *30th IEEE International Conference on Robot and Human Interactive Communication* (pp. 704-711). Vancouver, BC (online): IEEE.

[6] Jost, C., Le Pévédic, B., Belpaeme, T., Bethel, C., Chrysostomou, D., Crook, N., Grandgeorge, M. & Mirnig, N. (Eds.) (2020). *Human-Robot Interaction: Evaluation Methods and Their Standardization*, (Vol. 12). Cham, Switzerland: Springer Nature.

[7] Krämer, N. C., von der Pütten, A., & Eimler, S. (2012). Human-agent and human-robot interaction theory: similarities to and differences from human-human interaction (pp. 215-240). In M. Zacarias and J. V. de Oliveira (Eds.) *Human-computer interaction: The agency perspective*. Springer, Berlin, Heidelberg.

[8] Jackson, R. B., & Williams, T. (2021). A Theory of Social Agency for Human-Robot Interaction. *Frontiers in Robotics and AI*, 267.

[9] Jung, M., & Hinds, P. (2018). Robots in the wild: A time for more robust theories of human-robot interaction. *ACM Transactions on Human-Robot Interaction*, 7(1), Article 2, pp. 5.

[10] Wacker, J. G. (1998). A definition of theory: research guidelines for different theory-building research methods in operations management. *Journal of operations management*, 16(4), 361-385.