

Preface

This volume contains the Proceedings of *Robophilosophy 2020: Culturally Sustainable Social Robotics*, the fourth event in the biennial *Robophilosophy Conference Series*. The series presents interdisciplinary research in philosophy (and other Humanities) in and on social robotics. Past events in the series (*Robophilosophy 2014: Sociable Robots and the Future of Social Relations*, Aarhus; *Robophilosophy 2016: What Social Robots Can and Should Do*, Aarhus; *Robophilosophy 2018: Envisioning Robots in Society—Power, Politics, and Public Space*, Vienna) featured 60-100 research presentations and attracted 150-250 international participants. That these past conferences, as well as *Robophilosophy 2020*, are to date still the world's largest conferences in Humanities research in and on social robotics is problematically disproportional to the enormous projected economic significance of social robotics. However, participation numbers have been steadily rising and the research community of robophilosophy conferences is expanding. Moreover, supported by a new methodological reflectiveness in Human-Robot Interaction (HRI) research, it appears that, very slowly, those changes in the research landscape are taking place that motivated the institution of the *Robophilosophy Conferences* in the first place: if engineering products are to participate in human social interactions, new alliances must be established between engineering and SSH disciplines, with special emphasis on the humanities. The particular expertise of the humanities is the analysis of the *symbolic* and *normative* space of human interaction—what it means for individuals, communities, and societies to be engaged in this or that interaction with natural or technical systems, or with other human agents—and it is an expertise that social robotics and HRI research ultimately cannot do without. The realization of this circumstance may have been slower than one might have wished for—among both engineering *and* Humanities researchers—but as “human-centered AI and robotics” now begin to find explicit attention in research funding programs, the goal of producing “culturally sustainable” technology based on Humanities expertise should hopefully receive momentum.

These were the thoughts that went through the minds of the organizers of RP2020 in March 2020, at the onset of the COVID-19 pandemic, when we decided not to postpone the conference. While the dimension of the personal encounters in formal and informal discussions at a conference cannot be overstated, we felt that it was even more important to ensure that the unusually large number of submissions by younger researchers would receive a timely publication outlet. We decided that we would try to recapture some of the benefits of a live conference in creating a hybrid event—with live and pre-recorded content online, accessible during an extended period of time via an interactive webpage. We asked the speakers for the session papers to provide short video-recordings of their talks, which registered participants could comment on during a period of 10 days; these comments together with other questions were taken up in ten focused live online discussion sessions. Together with six live plenaries and five live workshops, moderated from a live conference studio, this large-scale research exchange filled the period from August 10-August 21, with live sessions occurring the last three days.

As an online event, the *Robophilosophy 2020* conference literally took place around the world. Due to a special advertising effort—which our plenary speakers generously

supported by collaborating with us on video teasers of their talks—close to 400 researchers from 29 countries around the globe participated. Colleagues in Europe, the US, Japan, New Zealand, Australia, the Philippines, China, and several countries in South America and the Baltics, joined in a debate about how to address the challenges of creating culturally sustainable social robotics. Thus, the COVID-19 pandemic, which likely will accelerate the global interest in social robotics, also forced researchers to adapt to a communicative format that reveals the concerns of robophilosophy as global concerns.

Which applications of social robotics (if any) could we rationally want? This may be the shortest formulation of the core question of robophilosophy. The term “robophilosophy” identifies an ongoing “fundamental systematic reconfiguration of philosophy in the face of artificial social agency” that involves three research dimensions—it is “philosophy *of, for, and by* social robotics”.¹ Robophilosophy is (i) the philosophical reflection of the socio-cultural and ethical impact of social robots; (ii) it is the employment of philosophical methods (conceptual and phenomenological analysis, formal theory construction, rational value discourse, etc.) for conceptual and methodological problems arising with artificial social agency; furthermore, (iii) it is experimental philosophy undertaken not merely with the familiar (quantitative, qualitative, experimental) methods of empirical research but also by construction (i.e. design and programming of physical and kinematic appearance and interactive capabilities).

Each robophilosophy conference articulates the core question from an angle that, in the perception of the local organizers, ties in with focal points of the current research discussion and public debate. The main agenda of RP2014 was to communicate the need for Humanities expertise in social robotics and HRI research. RP2016 served to delineate robophilosophy more clearly from roboethics and put the focus on the entanglement of theoretical, methodological, and practical-normative problems arising with social robotics. RP2018 stressed the larger socio-political implications and cultural dimension of the role of social robotics. The aim of RP2020 was to direct the challenge back to the research community in the Humanities—relative to recent developments in the research debate and a greater opening towards the Humanities, it seemed the right time to shift from critique to construction. Instead of criticizing omissions in HRI and social robotics research, we wanted to invite our colleagues to offer concrete proposals for *how* the Humanities can contribute to shaping a future where social robotics is guided by the goals of enhancing socio-cultural values rather than mere utilities.

After a decade of interdisciplinary research into social robotics and Human-Robot Interaction (HRI) we still lack a clear understanding and regulative directives for *how to ensure that social robotics will contribute to a community’s resources for human well-being*—to the practices in which members of a community experience justice, dignity, autonomy, privacy, security, authenticity, knowledge, freedom, beauty, friendship, sensitivity, empathy, compassion, creativity, and other socio-cultural core values, as these may be shared, or vary, across cultures. In the Call for Papers for RP2020, we invited philosophers and more broadly Humanities researchers to offer constructive answers to questions of method and procedure:

- Precisely *what*, in terms of conceptual tools and research methods, can Humanities researchers, who are trained in the analysis of the experiential complexity of human

¹ Seibt J. Robophilosophy. In: Posthuman Glossary. R. Braidotti, M. Hlavajova, editors. Bloomsbury; 2017. p. 390–4.

social interactions, contribute to the task of producing culturally sustainable applications of social robotics?

- Precisely *how* can Humanities research assist us in determining which socio-cultural values we wish to sustain or even to enhance?
- Precisely *how* can philosophers and other Humanities researchers assist engineers in exploring what interacting with ‘social’ robots will come to *mean* to us, as individuals and societies?
- And even more constructively, precisely *how* can we create cultural dynamics with or through social robots that will not impact our value landscape negatively? How could we design human-robot interactions in ways that will positively cultivate the values we, or people anywhere, care about?

Answers to these and related questions, by over 100 authors of 74 researcher contributions, are contained in these Proceedings.²

The systematic structure of the Proceedings deviates somewhat from the systematic structure of the conference, which was partly necessitated by practical issues (time zones). Here we have put conceptual, methodological, and design issues in front, in order to emphasize that the normative problems arising with social robotics in most cases cannot be addressed without clarifying conceptual issues beforehand or alongside.

While three of the five conference workshops advanced central themes in robophilosophy—sociality, moral standing, and trust—the remaining two workshops presented something novel. The workshop “Robots in Religious Contexts” might mark the beginning of a new research line—robotheology. The workshop “Think-and-Perform Tank”, on the other hand, introduced a new heuristic methodology for the development of culturally sustainable social robotics applications by crossing aesthetic and theoretical epistemologies in interactive improvisation between humans and machines.

The inclusion of art, as distinctive epistemic pathway to the ‘truth’ of human-robot interaction, has been a characteristic of all Robophilosophy Conferences, and for RP2020 we had invited the German theater ensemble “Meinhardt-Krauss” to perform their newest production “ELIZA—Uncanny Love” at the Music Hall in Aarhus. In order to include the symbolic trajectories of this play in some fashion, the artists agreed to produce a film version of selected scenes and in an “artist-audience” dialogue some insights and impressions could be shared. However, here more than elsewhere direct physical experience is decisive, and we are thus looking forward to the live performance of this play at the next conference in 2022.

To conclude with a look ahead, the tasks of robophilosophy cannot be fruitfully addressed from armchairs, ivory towers, or any reflective stance that isolates itself from the dynamics of praxis. As we develop artificial ‘social’ agents, the Humanities need to take on a new role and become pro-active, in order to help us to create technological futures worth living. There are currently two main strategies in robophilosophy.

On the one hand, some researchers engage in the wide-scope commentary of cultural criticism that reflects the role of social robotics in a larger cultural context—this is philosophy *of* social robotics. The audience of these reflections is typically society at large, and they aim to engender a shift away from pure profit maximization by changing *the minds* of individuals, and thereby changing practices. Let us call this the edification strategy.

² Videos of the plenary lectures are available at the Robophilosophy YouTube channel.

On the other hand, as documented by the majority of the contributions in these Proceedings, we see an increase in philosophy *for and by* social robotics: concrete proposals for new models and methods, presented by philosophers and other SSH researchers, for how to develop social interactions with robots in culturally sustainable ways. This may be called the optimistic-subversive strategy—it is the trust that by changing the conceptual tools and paradigms of production processes we can change the products and thereby change practices and minds in one step.

Only time can show which strategy will be more successful, and we probably need both. We hope, however, and are encouraged by the increasing number of participating non-philosophers, from the empirical sciences and from engineering, that the next robo-philosophy conferences will focus on strengthening the second strategy. In our view, the research actions we need in the future are not at the level of polite or antagonistic dialogue between technology development and the Humanities, but at the level of direct practical collaboration. Cultural sustainability is not something to be deduced from fixed premises, but something to co-develop in praxis.