



Robotics Aids for Character Building: More than Just Another Enabling Condition

Massimiliano L. Cappuccio¹ · Eduardo B. Sandoval² · Omar Mubin³ · Mohammad Obaid⁴ · Mari Velonaki²

Accepted: 26 January 2021 / Published online: 15 February 2021

© The Author(s), under exclusive licence to Springer Nature B.V. part of Springer Nature 2021

Can robotics play a moral role in the life of humans? This question has been systematically addressed by the main approaches to machine ethics, which are prevalently inspired by normative theories in the consequentialist and deontological traditions. These theories focus on the maximization of practical and psychological advantages for the users and the obligation to protect human dignity, respectively. In many circumstances, for example when designing autonomous vehicles, the careful implementation of these theories is indispensable to assure that the systems' behaviour aligns with the ethical intuitions that are shared by most people and enshrined in our systems of laws, such as the fundamental principles pertaining to common good or individual rights [7].

However, in addition to providing utilitarian benefits to the users and diligently imposing severe normative prescriptions, social robots can play a moral function also in another important sense: through repeated, interactive engagement in familiar scenarios, appropriately designed robots can help

the human users recognize and correct their bad habits, dedicate themselves to commendable activities, and incorporate values as spontaneous inclinations to think and behave virtuously. Like other forms of virtue cultivation, this process may lead, eventually, to fortify one's moral character and adopt a better way of living. This kind of moral betterment is central to virtue ethics, the normative theory that construes morality around the idea of living a praiseworthy lifestyle.

The assumption behind this idea is that true fulfilment and enduring happiness can only be accomplished through self-improvement and constant engagement in admirable activities. Specifically, developing one's good nature is possible by recognizing and correcting one's own flaws, following the inspiring moral example of others, and learning to embody their positive habits and qualities. It seems that today's robots are still unable to do any of these things for us, but they can already help us doing all of them better, if we decide that that is our goal. This is the key idea of what we call "virtuous robotics", an approach to machine ethics and social robots design inspired by the theory of virtues.

Virtue ethics has enjoyed a significant renaissance during the last few decades, thanks both to the philosophers who rediscovered its importance within the Western (e.g., Aristotelian) tradition and those who explored the links with its East-Asian (e.g., Confucian) homologous. During the last ten years, especially thanks to the work of philosophers like Mark Coeckelbergh, Rob Sparrow, and Shannon Vallor, virtue ethics became a prominent option in the field of technology ethics, as epitomised by Vallor's influential monograph *Technology and the Virtues* [2, 5, 6]. The foundational work conducted by these pioneers inspire in more than one way this special issue.

The approach to technology ethics inspired by virtue theory operates under the assumption that our technological practices are not value-neutral, but importantly contribute to shape our understanding of what is right and wrong. Moral values are always embedded, more or less implicitly, in the way technology is designed and adopted. By channelling

✉ Massimiliano L. Cappuccio
m.cappuccio@unsw.edu.au

Eduardo B. Sandoval
e.sandoval@unsw.edu.au

Omar Mubin
O.Mubin@westernsydney.edu.au

Mohammad Obaid
mobaid@chalmers.se

Mari Velonaki
mari.velonaki@unsw.edu.au

¹ School of Engineering and IT, University of New South Wales Canberra, Canberra, Australia

² School of Art and Design, University of New South Wales, Sydney, Australia

³ MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, Australia

⁴ Department of Computer Science and Engineering, Chalmers University of Technology, Gothenburg, Sweden

such values into everyday life, autonomous technologies can be leveraged to exert a positive or negative influence on the user's character. Our approach, that we feel close in spirit to the proposal advanced by Paul Dumouchel and Luisa Damiano in their book, *Living with Robots. Emerging Issues on the Psychological and Social Implications of Robotics* [3], particularly focuses on the moral role played by social robots and their transformative potential. Embodied humanoid agents have an unprecedented power to shape our conduct, our relationships, and our perception of what is ethically acceptable and desirable in the interpersonal sphere.

One important transformative aspect of our relationship with autonomous technology is that we often feel inclined to perceive social robots as moral patients, which raises the question—systematically addressed by David J. Gunkel in his recent book *Robot Rights* [4]—concerning whether or not robots deserve some form of moral consideration. Because they seem to claim for themselves the roles of both moral agents and moral patients, social robots—more than any other assistive technology—promise to transform not only the evaluative categories that we adopt in the legal and public policy domains, but also, more deeply and less obviously, the spirit of our customs and social norms.

Virtuous robotics reinterprets the groundwork established by our predecessors to put forward an entirely original approach to human–robot interaction. Our approach intends to offer a systematic response to the need, perceived more and more urgently by both makers and scholars, of a new human-centric normative framework to regulate autonomous assistive technologies without hindering the expression of their innovative potential.

Unfortunately, today's discussions in robot ethics are still narrowly focused on the moral risks that human–robot interaction could possibly bring to our society. To be clear: discussing the ethical risks associated with social robotics is indispensable because, without appropriate regulation and scrutiny, assistive technologies (either because inappropriately used or because loaded with misleading expectations) may eventually produce a tangible deterioration of the socio-ethical and political fabric of interpersonal relations. However, it would be impossible to assess the real transformative potential of social robotics without addressing the opposite, positive side of the story as well: by promoting a vigilant optimism, virtuous robotics can help us express a legitimate hope that technology will be capable to bring about authentic human development and opportunities of moral betterment.

Unlike other approaches to robot ethics, we are not simply advocating compliance with the ethical standards dictated by an abstract normative theory. That is why we will not limit ourselves to hope that assistive technologies were used to create the necessary material and intellectual pre-conditions for moral flourishing (for example, freeing people

from urgent needs and inequalities). We certainly share this hope, but we also believe that, compared to it, our project is more ambitious and more concrete at once: we envision that autonomous technologies, if considerably designed and sensibly integrated in our cultural forms of life, will soon be called to play not only the role of moral *enablers*, but also that of moral *facilitators* and *enhancers*. This means that social robots can and should be used to actively, concretely contribute to the development of their users' moral postures in practically significant scenarios, and not just to the fulfilment of some abstract rules or the formation of some generic ethical beliefs. If this hypothesis is correct, then a new generation of robots should be primarily conceived and specifically designed to promote moral flourishing in humans, aiming to make tomorrow's users wiser and gentler people, with more decent and realized lifestyles. This development is in principle possible if virtuous technologies are conceived to play an expansive and supportive role in the moral domain, analogously to how some assistive technologies are already successfully used to augment our resources and capabilities in the cognitive domain.

To give voice to this hope, but also to carefully explore the most efficacious strategies to pursue it, this special issue intends to address whether and how the responsible design and application of social robots could actively promote true moral awareness in humans and support their ethical decisional processes, considering both the case studies that favour such an optimistic expectation and the critical objections that encourage prudence.

We would like to acknowledge how our journey has crossed the path of several other studies and researchers in social-robotics, human–robot interaction, and technology ethics, importantly benefitting from these encounters. Some of our ideas about a virtuous approach to robotics were initially explored during a panel discussion held on 9 August 2017 at the Philosophy Department of University of Wollongong as part of the first UOW Social Robotics Workshop. This discussion, which primarily revolved on a paper published in the same year by Sparrow [5] in *International Journal of Social Robotics*, inspired Massimiliano Cappuccio, Anco Peeters and William McDonald [1] to respond with another paper on moral consideration for robots based on virtue and recognition, which was published in *Philosophy & Technology* in early 2019. Some of these ideas were further developed during a thematic session on “Robots and Virtue” (co-organized by Fady Alnajjar, Massimiliano Cappuccio, Friederike Eyssel and Mohamad Eid) held on 4–7 February 2018 at United Arab Emirates University and New York University Abu Dhabi. Then again, a seminal investigation on flourishing informed the second day of activities of the interdisciplinary workshop “Robots & AI in society”, co-organized by Omar Mubin and Massimiliano Cappuccio and held on 8–9 November 2018 at

Western Sydney University in collaboration with the Digital Humanities Research Group.

Summarizing the outcome of this journey, this special issue seeks to explore from a multidisciplinary perspective the idea that, if effectively programmed for pedagogical, training, and self-development purposes, robots could be used to cultivate virtues like compassion, honesty, and generosity, learn to manage negative emotions like anger and envy, and practice positive attitudes towards others, such as respect, care, and friendliness. The nine original papers selected for publication in this special issue investigate the philosophical, psychological, and technological/implementation dimensions of this hypothesis.

The paper that opens the collection is co-authored by the five guest editors of this special issue. Its aim is to provide a solid conceptual foundation to the “virtuous robotics” project, rooting this proposal into the principles of virtue theory and distinguishing it from other approaches to machine ethics. The paper explores functions, tasks, roles, and design principles that best suit the virtuous application of social robots and discusses a number of theoretical and practical objections.

The second contribution is authored by Australian technology ethicist Rob Sparrow. The well-argued doubts that he raises in this paper somehow temper the optimistic perspective presented in the previous one. Sparrow argues that a structural asymmetry informs our relationship with robots: while social robots can promote the cultivation of vice, they cannot help humans develop their virtues. The paper discusses the reasons and ethical implications of this asymmetry.

The third paper is authored by a technology ethicist working in the European philosophical tradition, Mark Coeckelbergh. Giving voice to a post-phenomenological perspective, his analyses highlight the need to further emphasize the relational dimension of moral character in human–robot interactions and the dependency of virtue on systems of cultural practices, critically discussing the absolute primacy incorrectly attributed by the Western ethical traditions to individual agency and rational deliberation.

The fourth paper, co-authored by Merel Keijsers, Hussain Kazmi, Friederike Eyssel, Christoph Bartneck is an empirical study in human–robot interaction informed by the rigorous methodology of social psychology. The study focuses on the psychological mechanisms underlying vandalism and abusive acts conducted against robots in public spaces. This study investigates how aggressive and punitive attitudes toward robots correlate to technology acceptance, desires to assert dominance, and the tendency to attribute intelligence and mentalistic features to robots.

The fifth contribution to the special issue¹ examines a well-known dilemma in the virtue ethical tradition—i.e. the philosophical question concerning why mistreating non-human animals is ethically unacceptable—and transposes it into the contemporary debate concerning the moral consideration for social robots. The four authors of the paper, Simon Coghlan, Frank Vetere, Jenny Waycott, and Barbara Barbosa Neves argue that robots may affect the moral development of children and, in general, the human responses to nonhuman animals, and consider the implications for design and policy.

In the sixth paper included in the collection, its authors, Anco Peeters and Pim Haselager discuss the psychological background and the moral implications of designing virtuous sex robots, that is social robots specifically conceived to engage in intimate relationships with humans. Is the existence of such robots, and the desire to utilise them to satisfy erotic needs, inherently vicious, or can these robots be developed with a moralising and educational intent, considering also the alleged deceptiveness of humanoid robots and the fact that they can be used instrumentally, without receiving consent from the robot itself?

Social virtues require the ability to empathise with others and regulate one’s own conduct to establish joint attention with them. But what happens when these abilities are disrupted and what can be done to recover them? The seventh paper, co-authored by Fady Alnajjar, Massimiliano Cappuccio, Abdulrahman Renawi, Omar Mubin, and Chu Kiong Loo, addresses this question, presenting the results of an experimental study on the beneficial impact played by robots to facilitate interactive engagement and maintain social attention in autistic children.

The eighth paper investigates a key anthropomorphic feature of social robots, i.e. their capability to communicate via facial expressions. The authors, Chris Chesher and Fiona Andreallo, explain how robot faciality plays not only an aesthetic function but also an ethical one, soliciting moral awareness and self-regulation in the human users. Unlike face-less technologies, social robots can rely on their face and gaze to make humans feel recognized as objects of moral evaluation and hence as moral agents who carry the burden of ethical responsibility.

The ninth paper offers decisive empirical evidence that robots can be efficaciously deployed in public contexts to encourage ethical behaviours in casual interactants, educating them to respect deontological norms and fulfill their civic duties. The study conducted by Berardina De Carolis,

¹ Note that, although this paper was originally written to be included in our thematic collection on Virtuous Robotics, it has already been published by *International Journal of Social Robotics*. The paper is not reproduced in this special issue, but readers can find it in the December 2019 issue.

Francesca D’Errico, Nicola Macchiarulo, and Veronica Rosano successfully tested robot-mediated playing as a method to promote ecological awareness among children, who acquired environment-friendly recycling habits as a result.

These papers helped us understand whether and how social robots (especially those explicitly programmed to act as coaches, companions, or training assistants) can inculcate new virtues or assist humans in developing good habits, accompanying their personal search for a morally aware and more balanced way of living. We hope that new studies will be designed and conducted to explore other applications of virtuous robotics in an even greater variety of scenarios and tasks, further clarifying—through the use of socially interactive robots—the cognitive and social mechanisms that underpin the formation of character and the cultivation of virtue in adults and children.

References

1. Cappuccio ML, Peeters A, McDonald W (2019) Sympathy for dolores: moral consideration for robots based on virtue and recognition. *Philos Technol* 2019:1–23
2. Coeckelbergh M (2018) Why care about robots? Empathy, moral standing, and the language of suffering *Kairos*. *J Philos Sci Sciendo* 20(1):141–158
3. Dumouchel P, Damiano L (2017) *Living with robots*. Harvard University Press, Cambridge
4. Gunkel DJ (2018) *Robot rights*. The MIT Press, Cambridge
5. Sparrow R (2017) Robots, rape, and representation. *Int J Soc Robot* 9(4):465–477
6. Vallor S (2016) *Technology and the virtues: a philosophical guide to a future worth wanting*. Oxford University Press, Oxford
7. Wallach W, Allen C (2009) *Moral machines. Teaching robots right from wrong*. Oxford University Press, Oxford

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Massimiliano L. Cappuccio is Senior Researcher at the University of New South Wales, where he serves as Deputy Director of the Values in Defense and Security Technology Group. He also holds a secondary affiliation with UAE University, the national university of the United Arab Emirates, where he had been Associate professor of Cognitive science and director of the Cog Sci Lab (August 2011–December 2018). As a cognitive philosopher and a technology ethicist, his research is interdisciplinary and aims to integrate phenomenological analyses, empirical experimentation, and synthetic modelling. His work focuses on theoretical and applied issues in human performance (skill acquisition and skill disruption), human–robot interaction (social and assistive robotics), philosophy and theory of artificial intelligence (frame problem, computational theory of mind). His philosophical perspective attempts to integrate embodied, enactive and extended approaches to social cognition, virtue ethics, and complexity theory. He conducts an intense activity as organizer of academic events, including interdisciplinary workshops, research seminar series, and international conferences (like the TEMPER workshop on Training, Enhancement, and

Military Performance at ADFA, and the annual Joint UAE Symposium on Social Robotics in Abu Dhabi).

Eduardo B. Sandoval is Scientia Associate Lecturer at the University of New South Wales (UNSW) since 2018. Dr. Sandoval has an interest in the design of machines that are capable of meaningful social interactions with humans. Dr. Sandoval’s work incorporates insights from behavioural economics and social psychology to explore different approaches in social robotics. His pioneering work has evolved around the topics of Reciprocity in Human Robot Interaction (HRI), Robot Design, and Robot Games. Similarly, he has explored the use of robots in education and healthcare and future addiction to robots in the academia and industry. Dr. Sandoval has produced thirty-four peer-review publications in relevant venues related to HRI, a robot design register in 2019, and his work in Human–Robot Interaction has circulated in international media. Similarly, he has been appointed as co-chair, guest editor, and reviewer for several conferences and journals related to HRI, and Social Robotics and has collaborated closely with industry in the design and development of robot prototypes. In 2014, The National Autonomous University of Mexico awarded Dr. Sandoval with the Medal Alfonso Caso as the best student of the Postgraduate Program of Industrial Design at the National Autonomous University of Mexico. In 2016, Dr. Sandoval obtained his Ph.D. in Human Interface Technology at the HITLab NZ, University of Canterbury, New Zealand. His future work involves multidisciplinary research in social robotics as robot games and no-visual natural interfaces aiming future human’s well-being.

Omar Mubin is a Senior Lecturer in Human Computer Interaction at the School of Computer, Data and Mathematical Sciences at Western Sydney University, Australia. Dr. Mubin’s primary research interests are human robot interaction and human-agent interaction. Specifically, he studies social robotics and their applications and consequently interaction with humans in education, public spaces and information dissemination scenarios. He also has a keen interest in the application of data analytics to the areas of Scientometrics, ICT and Public Health. His current h-index according to Google Scholar is 19 with a more than 1700 citations. Dr. Mubin is involved in teaching and supervising (undergrad and postgrad) students in the broader area of Human Computer Interaction, Mobile Computing and Health Informatics.

Mohammad Obaid is an Associate Professor and Head of the Interaction Design Unit at the Department of Computer Science and Engineering, Chalmers University of Technology (Sweden). Dr. Obaid received his B.Sc., M.Sc. (First Class Honors) and Ph.D. degrees in Computer Science from the University of Canterbury, Christchurch, New Zealand, in 2004, 2007 and 2011, respectively. In April 2018, he received his Docent degree (Associate Prof. rank) from Uppsala University (Sweden). He worked at several international research centres including the Human Centered Multimedia Lab (HCM Lab), University of Augsburg (Germany), the Human Interface Technology Lab New Zealand (HITLab NZ), University of Canterbury (New Zealand), the t2i Lab, Chalmers University of Technology (Sweden), and the Social Robotics Lab, Department of Information Technology, Uppsala University (Sweden). Dr. Obaid is one of the founders of the Applied Robotics Group at the Interaction Design Division at Chalmers University of Technology (established in 2015). He is the (co-)author of over 70 publications within the areas of his research interests on Human–Robot Interaction and Human–Computer Interaction.

Mari Velonaki’s research is situated in the multi-disciplinary field of Social Robotics. She holds a Ph.D. in Experimental Interface Design (UNSW 2003). Velonaki began working as a media artist/researcher in the field of responsive environments and interactive interface design in 1997. She pioneered experimental interfaces that incorporate movement, speech, touch, breath, electrostatic charge, artificial vision and

robotics, allowing for the development of haptic and immersive relationships between participants and interactive agents. In 2014 she was voted by Robohub—a large robotics community of researchers, educators and business- as one of the world’s 25 women in robotics you need to know about. She is the founder and director of the Creative

Robotics Lab (Art & Design UNSW) and the founder and director of the National Facility for Human Robot Interaction Research (UNSW, USYD, UTS, St Vincent’s Hospital). Mari is a Research Leader at the UNSW Ageing Futures Institute.