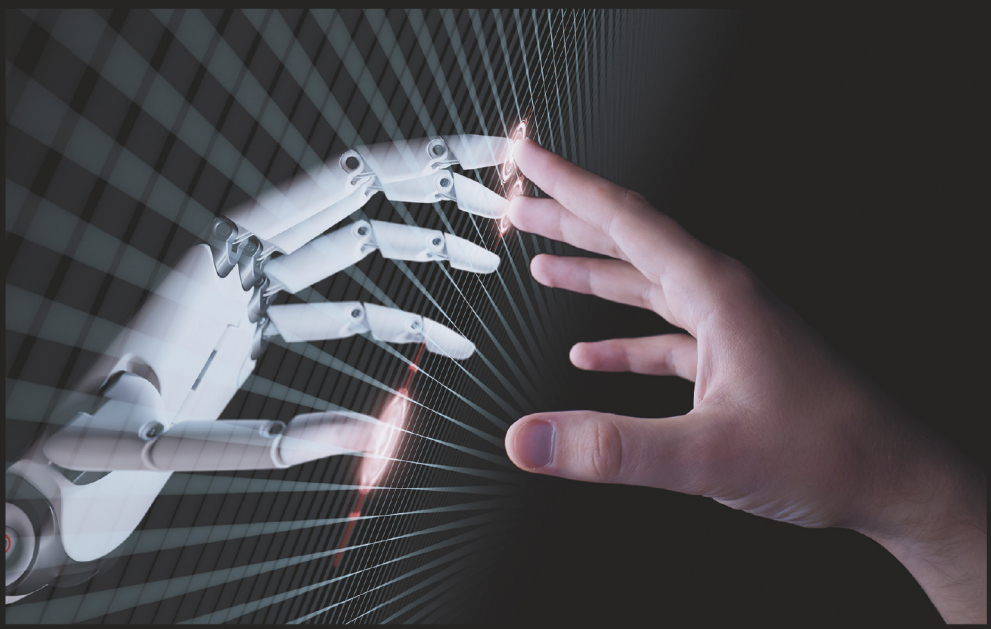
GIUSEPPE RIVA ANTONELLA MARCHETTI (EDS.)

HUMANE ROBOTICS

A MULTIDISCIPLINARY APPROACH TOWARDS THE DEVELOPMENT OF HUMANE-CENTERED TECHNOLOGIES



VP VITA E PENSIERO | RICERCHE

HUMANE ROBOTICS

GIUSEPPE RIVA
ANTONELLA MARCHETTI (EDS.)

HUMANE ROBOTICS

A MULTIDISCIPLINARY APPROACH TOWARDS THE DEVELOPMENT OF HUMANE-CENTERED TECHNOLOGIES

VITA E PENSIERO | RICERCHE

Questa ricerca e la sua pubblicazione sono state finanziate dall'Università Cattolica nell'ambito dei suoi programmi di promozione e diffusione della ricerca scientifica (linea D3.2. anno 2018, "Human-Robot Confluence" project).

www.vitaepensiero.it

ISBN edizione cartacea: 978-88-343-4618-1
ISBN edizione digitale (formato PDF): 978-88-343-4619-8

Copertina: Studio grafico Andrea Musso

INDICE

Preface

Università Cattolica del Sacro Cuore has launched numerous research projects dedicated to studying the relationship between human and technology. The importance of this connection has become increasingly pressing in the public debate of recent years; however, our University can lay claim to a long tradition of studies, the origins of which go back several decades. One of the most immediate memories is linked to the figure of Father Roberto Busa (1913-2011), founder of computational linguistics and our teacher. His research is still a point of reference for many scholars in the field, given that where the most evident advances in Artificial Intelligence (AI) systems have been made is precisely the field which Father Busa pioneered. Thanks to his generosity, we are proud to have made available to students and lecturers his library and personal archives, which he donated to us shortly before his death and which today represent a valuable source for new investigations.

This historical reference is intended to underline how current analyses of the relationship between technologies and human experience are framed within a scientific tradition to which our University has always paid particular attention. The establishment of the Humane Technology Lab (HTLab), a laboratory that promotes and enhances research activities devoted to technologies and the various dimensions of human experience, is the most recent evidence of this commitment. This publication is in fact one of the first fruits of the HTLab, conceived and brought to fruition by a team of high-profile scholars – coordinated by Giuseppe Riva and Antonella Marchetti – to whom I am particularly grateful for their work.

Reflecting on Humane Robotics, the authors of the essays published in the following pages demonstrate that a multidisciplinary approach is the most suitable to study such a complex topic. Individual knowledge, whether technical, legal, psychological, social or political, is not sufficient; a dialogue between different approaches is necessary to understand the radical nature of the challenge that the latest technologies bring to human life. As Pope Francesco reminds us, "A science which would offer solutions to the great issues would necessarily have to take

into account the data generated by other fields of knowledge", because the "fragmentation of knowledge proves helpful for concrete applications, and yet it often leads to a loss of appreciation for the whole, for the relationships between things, and for the broader horizon" (*Laudato si'*, no. 110). The most evident merit of this research is precisely that of providing a 'broad horizon' to the topic of Humane Robotics, capable of exalting one of the characteristics of our University, which is precisely that of carrying out research capable of systematically involving scholars from different disciplines and bringing it to the attention of a national and international public.

The underlying reason for this work is the need to develop a 'human' reflection on Artificial Intelligence, since the great progress it has brought about raises questions and is, in some ways, disturbing. Research has now reached previously unimaginable levels: while robots were once considered tools to replace humans in certain activities, they now seem to have become 'subjects' that act like humans, are physically similar to them and are (perhaps) in a position to think like them. Experts explain that the idea – or fear – of the assimilation of the processes of which machines are capable to any form of thinking is technologically and functionally improper and hasty, but the widespread perception is undoubtedly of a potential imminent substitution of machines in certain activities (and jobs) today entrusted to human. The radical nature of such a change is evident to those who have experienced first-hand the transition to a fully computerised and technology-dependent society. But are we sure that our current (and, even more so, our future) students are fully aware of this revolution? From this point of view, I hope that the results presented here will further stimulate the interest and curiosity of the new generations.

On the other hand, the central question posed by the development of Artificial Intelligence, and in particular the creation of intelligent robots, is indeed crucial: can we accept that a human can have an *other* who thinks and acts like him? This question encompasses many others, and among the four sections into which the book is divided, it is possible to find an interesting discussion on the consequences arising from this question, which of course does not exhaust the formulation of new hypotheses of study. Indeed, new questions and new doubts will arise, because at the heart of this reflection is the nature of human and his power to transform and control what is *other*.

It is perhaps no coincidence that one of the five guiding principles of the *Strategic Programme on Artificial Intelligence 2022-2024*, recently drawn up by the Italian government, states that Italian Artificial Intelligence will be anthropocentric, reliable, and sustainable. In other words, it will have to be implemented responsibly, transparently, and, above all, in a

human-centred way. As far as universities are concerned, this undoubtedly entails a commitment to strengthening skills related to these issues, increasing dedicated training courses, and taking steps to train and attract researchers in this field of study. But it is perhaps even more challenging to understand how, through which practices, according to which aims, Artificial Intelligence will have to be 'human-centred'.

The present research proposes a first reasoned answer to this question, thus continuing the commitment of Università Cattolica del Sacro Cuore to understanding the transformations of the ancient relationship between human and technology, its challenges and ambiguities. At the same time, it opens up further lines of enquiry, which I hope will be developed in the research planned for the coming years.

Franco Anelli
Rector of Università Cattolica del Sacro Cuore

INTRODUCTION

The Search for Human Robot-Confluence

G. Riva, A. Marchetti

ABSTRACT

The book brings together knowledge from multiple disciplines – mechatronics and computer science, psychology and neuroscience, philosophy and ethics, medicine, sociology, anthropology, economics, law and education among others – with respect to human-robot confluence (HRC) in the application of robots in everyday life, including assistive and rehabilitation robotics. It covers a wide range of topics related to human-robot confluence, drawing on various theories, methodologies, technologies, and empirical and experimental studies.
Moreover, in its final section, the book discusses the main challenges that we are facing in the search for humane technology. The final goal is to support researchers and developers in creating technologies that not only mimic human communication and cognition but are genuinely 'humane': accessible, sympathetic, generous, compassionate, and forbearing.

The Search for Human-Robot Confluence

The new humanoid robots not only perform tasks, but can also engage in interactions and social relationships with other robots and humans. In particular, the increasing use of humanoid robots is influencing many daily contexts – cooperative work, assistive living, monitoring, security, education and entertainment – generating frequent human-robot interactions in unstructured environments (Riva & Wiederhold, 2021).

This diffusion of humanoid robots – robots with a physical structure reminiscent of the human body – that are endowed with decision-making abilities and capable of externalizing and generating emotions is opening a new line of research aimed primarily at understanding the dynamics of social interactions generated by the encounters between ro-

---

The work contained in this chapter is an extension of the article originally published as Riva, G. & Wiederhold, B.K. (2021). Human-robot confluence: Toward a humane robotics. *Cyberpsychology, Behaviour, and Social Networking, 24*(5), 291-293.

bots and humans (Marchetti et al. 2018; Sciutti, Mara, Tagliasco, & Sandini, 2018).

As underlined by Vignolo and colleagus (2017, p. 1), "The success of the integration of robots in our everyday life is then subordinated to the acceptance of these novel tools by the population. The level of comfort and safety experienced by the users during the interaction plays a fundamental role in this process".

However, this process is not easy. As underlined by different authors, humanization (Giger et al., 2019) – mimicking human appearance and behavior, including humanlike cognitive and emotional states – is not enough. Many individuals who have not yet experienced direct contact with them tend to consider humanoid robots a possible threat, and this negative attitude impacts on their intention to interact with them or not (Piçarra & Giger, 2018). However, individuals who have interacted directly with humanoid robots also frequently report discomfort "and a sense of eeriness and revulsion" (feelings of 'uncanniness'; Gray & Wegner, 2012).

Apparently, to be accepted by society, robots must demonstrate that they can participate in a genuine social experience (Gray & Wegner, 2012), that they can 'understand' people and adapt themselves to complex real-life social environments. In other words, they must be 'humane': accessible, sympathetic, generous, compassionate and forbearing.

Recently, Sandini and Sciutti (2018) suggested a possible agenda for achieving humane robots, according to which they should:

1) gain intuition, becoming partners rather than just sophisticated tools;

2) think beyond real time;

3) use an anthropomorphic imagination for human-robot interaction.

A further point has been suggested by Leveringhaus (2018): an ethical framework that makes a commitment to human rights, human dignity and responsibility a central priority for developers and researchers working with humanoid robots.

This book is the result of the work of Università Cattolica del Sacro Cuore's Humane Technology Lab (HTLab – www.humanetechnology.eu), whose main objective is to consider the impact of new technologies on the various dimensions of human experience, from both a scientific and cultural perspective. The HTLab has gathered knowledge from different disciplines – psychology and neuroscience, philosophy, medicine, sociology, anthropology, mechatronics and computer science, economics, law and education – with respect to human-robot confluence (HRC) in the application of robots in everyday life (Gaggioli et al., 2016), in-

cluding assistive and rehabilitation contexts in which the human with whom robots interact, of all ages and with a range of medical conditions, must be considered (Marchetti et al., 2020). Using these multidisciplinary insights, the book covers a wide range of topics related to human-robot confluence, drawing on theories, methodologies, technologies, and empirical and experimental studies.

HRC has emerged over the last decade as a branch of the developing field of human-computer confluence (Gaggioli et al., 2016). HRC specialists are conducting fundamental and strategic research into ways of basing the emerging symbiotic relationship between humans and computers on radically new forms of interaction, sensing, perception and understanding.

A key aspect of human-computer confluence that is also relevant to robotic research is its potential for transforming human experience in terms of bending, breaking, and blurring the barriers between the real, virtual and augmented human. In robotics, HRC is already exploring these boundaries and asking questions like 'Can we seamlessly move between the human and the robot?' and 'Is an interaction with a humanoid robot equivalent to one with a human?'. From this perspective, HRC can be considered a tool for exploring key topics, such as quality of interaction, functionality, ethics and effectiveness, that require collaboration between multiple disciplines. To reflect this, the book is divided into sections based on four different themes.

The first section, *From Human-Robot Interaction to Human-Robot Experience. A Theoretical Framework*, discusses the theories and conceptual tools capable of driving the exploration of human-robot confluence.

The first chapter by Gaggioli and colleagues discusses the existing limitations of humanoid robots, which emerge when robots are faced with real-life contexts and activities occurring over long periods. In their view, it happens because collaboration is a complex *relational* process that entails mutual understanding and reciprocal adaptation. To overcome this issue, they suggest a change of paradigm, shifting from 'human-robot interaction' to 'human-robot shared experience'. In their view, HCI research should indeed focus on the emergence of a shared experiential space between humans and robots. On the one hand, this requires the introduction and use of new concepts such as co-adaptation, intersubjectivity and individual difference. On the other, it implies a significant change in current mainstream design approaches, which are still focused on the functional dimension of human-robot interaction.

The second chapter, by Fox and Gambino, challenges the classical vision of many human-robot interaction (HRI) studies that apply to this field the rules and theories of research on interpersonal interaction.

They argue that the starting point should be our knowledge of personal relationships, and this chapter presents predominant interpersonal theories, the primary claims of which can serve as the basis for our understanding of human relationship development (social exchange theories including resource theory, interdependence theory, equity theory and social-penetration theory). Moreover, they discuss whether interpersonal theories are viable frameworks for studying HRI and human-robot relationships, given their theoretical assumptions and claims. The chapter closes by providing suggestions for researchers and designers, including alternatives to equating human-robot relationships to human-human relationships.

The third chapter by De Florio, Chiffi and Fossa introduces the concept of autonomy, which is critical for the theoretical understanding of both robots and complex technological artifacts. As automation progresses, new technologies exhibit forms of behavior that appear to be autonomous, thus adding new layers of meaning to the concept and, at the same time, introducing the possibility of comparing human and artificial forms of autonomy. In this light, the chapter discusses the difference between two concepts of autonomy: autonomy of performance and autonomy of processes. First, functional autonomy is described as autonomy of performance, according to which an artificial agent can be said to be autonomous if and only if it is able to perform a task without requiring the intervention of other agents. Then, functional autonomy is described as autonomy of process, according to which an artificial agent is autonomous if and only if it can select the procedure for accomplishing a goal from among a set of alternatives.

In one respect, these concepts make it possible to differentiate the autonomy of a robotic arm from the autonomy of a highly complex machine, such as the rovers used to explore planets; in another, they lay the groundwork for future investigations into the ethics of artificial agents.

The final chapter of this section, by Mazzola and colleagues, suggests addressing the humane approach to robotics through the concept of experience. As humans, we can only achieve the authentic comprehension of another subject through our own embodied Self: we comprehend others starting from ourselves – an ability infants develop over the years. Attempts to provide robots with a primitive sense of Self and of being distinct from others have leveraged machine-learning techniques and used bodily interactions with the surrounding area as a starting point. To continue in this direction, the authors believe that cognitive robotics should receive inspiration from the human way of experiencing the world and others. The concept of an embodied Self as a pivot of experience is cardinal and provides evidence for the constitutional relationship between what is experienced and the subject of experience. Final-

ly, the possibility of developing an experiencing robot raises questions about the nature of human experience and the potential impact of such technologies.

The second section, *Exploring Human-Robot Interaction. The Influence of Cognitive and Affective Processes*, analyzes the characteristics of human-robot interaction, exploring the critical role that cognitive and emotional factors play in it.

The first chapter, by Jung and colleagues, used the 'uncanny valley' model to analyze affective and cognitive responses to service humanoids. In particular, the paper focuses on the effect of affective responses on trust – considered a critical cognitive factor in technology adoption – in two service contexts characterized by different levels of expertise: hotel reception (low expertise) and tutoring (high expertise). The results suggest that affective and cognitive responses are more positive for the high-expertise humanoids than the low-expertise ones. Moreover, the 'uncanny valley' effect is attenuated by the level of trust placed in robots. For this reason, people's attitudes are less influenced by humanoids' peripheral cues (e.g., appearance) when the tasks they are performing require higher levels of expertise. By providing a richer understanding of humans' affective and cognitive reactions to humanoids, these findings ultimately contribute to research on user adoption of service robots.

In the second chapter, Manzi and colleagues describe the development and testing of the 'Scale for Robotic Needs', used to explore the different expectations that people have of humanoid robots. Using a latent profile analysis, the chapter describes five profiles of expectation that can be placed along a continuum of robot humanization: from a group who consider robots to be no more than technological tools at the service of humans (i.e., mechanical properties) to a group that anticipates that robots will be part of our society soon (i.e., Self-determination). The study also suggests that negative attitudes toward robots are strongly related to people's expectations.

The third chapter, by Zhu and Deng, describes two studies used to investigate the effects of social anxiety on the adoption of robotic training partners among university students. The first study confirmed that university students with higher social anxiety are more likely to choose robotic training partners than human training partners. The second study underlined the mediating role of the sense of relaxation, suggesting that training robots can improve quality of life for socially anxious people.

In the fourth chapter, Manzi and colleagues explore whether robot gaze has similar effects to human gaze in infants. Specifically, the study presented used eye-tracking to explore the gaze of infants as they watched four video clips, where either a human or a humanoid robot performed

an action on a target. The agent's gaze was either turned towards the target (congruent) or away from it (incongruent). The results suggest the presence, in infants, of two distinct levels of gaze-following mechanisms: one recognizing the other as a potential interactive partner, the second recognizing the partner's agency. In the study, infants recognized the robot as a potential interactive partner, whereas they ascribed agency more readily to the human, thus suggesting that the process of generalization of gazing behavior towards non-humans is not immediate.

In the fifth chapter of this section, Hanoch and colleagues explored a classic topic of social psychology: peer pressure. A large body of evidence has shown that peer pressure can impact human risk-taking behavior, but we do not yet know if the presence of robots can have a similar impact or not. Therefore, the study evaluated participants' risk-taking behavior when alone, when in the presence of a silent robot and when in the presence of a robot that actively encouraged risk-taking behavior. The results revealed that participants who were encouraged by the robot did take more risks, but the presence of a silent robot did not entice participants to exhibit more risk-taking behavior.

The sixth chapter, by Manzi and colleagues, examined children's attribution of mental states to two humanoid robots, NAO and Robovie, which differ in their level of anthropomorphism: the NAO robot presents more human-like characteristics than the Robovie robot, whose physical features appear more mechanical. The results showed that five-year-olds have a greater tendency to anthropomorphize robots than older children, regardless of the type of robot. Moreover, the findings revealed that, although children aged seven and nine attributed some human-like mental attributes to both robots to a certain degree, they attributed higher mental qualities to NAO than to Robovie compared to younger children. These findings have important implications for the design of robots, which must also consider users' target age, as well as for the generalizability of research findings, which are commonly associated with the use of specific types of robots.

Nijssen and colleagues, in the seventh chapter of the section, provide an answer to a critical question for HCI: do we take a robot's needs into account during a common task? Using two experiments, they investigated whether individuals take the needs of a robotic task partner into account to the same extent as they do for a human task partner, and whether this was modulated by the degree to which participants' anthropomorphized said robot. The results suggest that humanizing a task partner does indeed increase our tendency to take someone else's needs into account in a social decision-making task. However, this effect was only found for a human task partner, not for a robot.

In the final chapter of the section, Manzi and colleagues examined

the different psychological effects generated by two commercial humanoid robots: NAO and Pepper. Their study assessed variability in the attribution of mental states, expectations regarding the sophistication of the robot, and negative attitudes after observing a real interaction with a human (an experimenter). The results suggest that both the observation of interaction and the physical appearance of the robot affect the attribution of mental states, with a greater attribution of mental states to Pepper than to NAO. People's expectations, however, were influenced by the interaction, regardless of the type of robot. Finally, negative attitudes were found to be independent of both the interaction and the type of robot.

The third section, *Field Research in Human-Robot Interaction*, presents and discusses the use of robotics in different fields of application, from medicine to agriculture.

The first chapter, by Valentini, D'Oria and Cesario, explored the use of artificial intelligence (AI) and AI-guided robotics in diagnostics and therapeutics. In particular, the chapter discusses four possible applications: AI/robot-assisted surgery and radiotherapy for mini-invasive treatments; AI algorithms for outcome prediction, assessment and patient enrollment in clinical trials; process-mining algorithms for pathways and clinical-trial appropriateness, and patient support programs for patient journey continuity. The chapter also explores the ethical and legal implications of directly involving AI in patient care and the unresolved challenges regarding validation, regulation and assessment issues. To this end, it is necessary both to create guidelines for the integration and correct use of AI in diagnostics and to further explore and improve human-machine interaction and acceptance.

Maggioni and Rossignoli, in the second chapter of this section, survey the main contributions in behavioral economics literature on the study of interactions between human and 'artificial' agents. While human-computer interactions have been extensively studied in the last decade, artificial intelligence and humanoid robots have been a focus of behavioral economics only very recently. The experimental evidence surveyed in the chapter shows that interactions with artificial agents trigger less of an emotional response in subjects, while subjects tend to extend similar attributes that they attach to human beings to humanoid robots, provided the robots are able to display emotion, empathy and effective and context-related communication.

In the third chapter, Manzi and colleagues discuss the results of different behavioral studies in the field of human-robot interaction, showing that robots are perceived as plausible human partners in various contexts. Moreover, studies combining neuroscience, cognitive science

and robotics generally suggest that our brain responds quite similarly when stimuli involve a human and a robotic agent, as long as the robot's visible behavior (i.e., movements and emotional expressions) resembles the human's, in which case motor and emotional resonance mechanisms are activated. However, the chapter identifies the limits of these claims, suggesting that our cognitive and control processes prevail over the tendency to anthropomorphize the robots after experiencing their limits in real interactions.

The fourth chapter, by Tosoni, Mascheroni and Colombo, presents a sociologically-inspired media study on social robotics. Specifically, the study investigated the relationships between humans and media-machines in natural settings, drawing on two cases of human-robot communication: the deployment of Pepper at Bologna Airport and at Università Cattolica del Sacro Cuore. The preliminary results demonstrate that interaction with social robots, far from being 'natural,' is prefigured and disciplined both by the robot's programming, configuration, and scripts, and by how humans are instructed to assume a defined – and limited – interactional role.

The final chapter of the section, by Gatti and colleagues, discusses how humans evaluate different robotic solutions that perform selective *vs.* non-selective and high-risk *vs.* low-risk operations in viticulture. In general, individuals prefer the use of humans over robots to perform selective agricultural activities, thus possibly reflecting the greater reliability associated with humans in making decisions and performing selective tasks. Nevertheless, people prefer the use of humanoid robots in selective activities requiring more complex decisions. In particular, there is a preference for the use of humanoid robots for performing selective tasks in both the safety and quality domains. Finally, autonomous vehicles are preferred to humanoid robots due to the increased productivity and reduced risk associated with the former.

The final section, *Towards a Humane Technology. Challenges and Perspectives*, shifts the focus to the global challenges that we are facing in the search for a humane technology.

Riva, Wiederhold and Mantovani, in the first chapter of this section, use recent neuroscience research findings to explore how distance learning and smart working are impacting the following three pillars at the core of school and office experiences: a) the learning/work happens in a dedicated physical place; b) the learning/work is carried out under the supervision of a boss/professor; and c) the learning/work is distributed between team members/classmates. The analysis presented suggests that the use of technology has a significant impact on many identity and cognitive processes, including social and professional iden-

tity, leadership, intuition, mentoring, and creativity. In conclusion, simply moving typical office and learning processes to a videoconferencing platform, as happened in many contexts during the COVID-19 pandemic, can erode corporate cultures and school communities in the long term.

In the second chapter, Rivoltella presents and discusses the so-called 'fourth wave' of media development, characterized by the pervasive presence of media and the central role of data and platforms. The pervasiveness of this mechanism is producing a new form of capitalism – a 'surveillance capitalism' – within which everyone is traceable, and value is equated to information. This situation affects our way of thinking about media education: today's media education is data literacy, above all. In this light, the chapter identifies different tools for critically analyzing texts and exercising active citizenship. The result of this work enables the identification of spaces within educational contexts in which we can raise people's data literacy, helping them live in an increasingly critical and autonomous way.

The third chapter, by Malavasi and Giovanazzi, suggests the need for a critical and conscious approach to machines – the 'pedagogy of artificial intelligence' – as an educational challenge for the development of civilizations, based on a vision of humanism centered on its engagement with the potential and limits of technological processes. This topic is also discussed in relation to the World Expo in Dubai (1st October 2021 to 31st March 2022), which has adopted the emblematic issues of technological innovation and sustainability as its main theme: 'Connecting Minds, Creating the Future'. From this perspective, an 'authentic' culture of scientific and technological research is based on the educational processes and creative dynamism that see innovation as an experiential space for generating fields of action for the sustainable and integrated development of humanity.

The fourth chapter, by Riva, Mantovani and Wiederhold, discusses the potential of positive technology to augment and enhance existing strategies for generating psychological well-being during the COVID-19 pandemic. Different positive technologies – mHealth and smartphone apps, stand-alone and social virtual reality, video games, exergames and social technologies – have the potential to enhance various key features of our personal experience – affective quality, engagement/actualization and connectedness – that are impacted by the pandemic and its social and economic effects. In this regard, positive technologies can be extremely useful for reducing the psychological burden of the pandemic and helping individuals to flourish even in difficult and complex times.

In the final chapter of this section, Della Morte suggests that Big Tech

companies are asserting themselves as centers of power that exercise public functions. In fact, presiding unchallenged over entire areas of cyberspace, every time they modify their computer code, they generate norms. One possible example are the rules governing access to digital data in the event of death. Establishing how our digital identities operate after our passing means establishing a kind of inheritance law, potentially applicable to billions of individuals. These transformations are a matter of concern for jurists, who are identifying a transfer of regulatory powers from states and international actors to those in the private sector. However, law concerns principles and values, not numbers, and the issue at stake is precisely how to reconcile the regulatory function of law with the rationale that increasingly underlies policies based on data computation. What we need is to identify a suitable set of principles for building a compass to navigate the deep and treacherous waters of cyberspace.

The final *Appendix* includes two documents that acknowledge the necessity of clear guidance for the future of the next generations, suggesting that no unconditional or emphatic acceptance of Artificial Intelligence and Robotics is possible.

The first, *Robotization of Life*, has been produced in 2019 by an *ad hoc* working group on robotization established by the Commission of the Bishops' Conferences of the European Union (COMECE). Led by professor Antonio Autiero and enriched by diverse contributions of experts in theology, philosophy, law and engineering, the COMECE working group analyzed the impacts of robotization on the human person and on society as a whole and elaborated its reflection as an ethical step which can shape community life in our complex and globalized society in which actors are increasingly interconnected. The document reaffirms the primacy of the human, on the basis of the recognition of the human dignity of each person.

The second, *Rome Call for AI Ethics*, is a document signed by the Pontifical Academy for Life, Microsoft, IBM, FAO and the Italian Ministry of Innovation in Rome on February 28th, 2020, to promote an ethical approach to Artificial Intelligence (AI). Its goal is to generate a sense of shared responsibility among international organizations, governments, institutions and the private sector in order to keep humankind at the center of digital innovation and technological progress. Paving the way for a new 'algorethics,' this document foresees the development of an artificial intelligence based on these cornerstone principles: AI serves every person and humanity as a whole; AI respects the dignity of the human person, so that every individual can benefit from the advances of techno-

logy; and AI does not have either greater profit or the gradual replacement of people in the workplace as its sole goal. Individuals and organizations interested in endorsing the Rome Call for AI ethics can find the relevant information here: https://www.romecall.org/joinus/.

In conclusion, the contents of this book constitute a sound foundation and rationale for future research aimed at exploring both human-computer confluence and human-robot confluence in the application of robots in everyday life. In particular, it provides strong preliminary evidence to justify future research into developing a new generation of humanoid robots that can acquire and demonstrate their participation in real social experiences. The challenge for research and developers over the next five to ten years is to design and develop technologies that can 'understand' people, and to build a shared communicative and relational experience (Riva & Wiederhold, 2021). Only in this way will it be possible to experience technologies that are 'humane' – accessible, sympathetic, compassionate and forbearing – and can truly support their human counterparts.

References

Gaggioli, A., Ferscha, A., Riva, G., Dunne, S., & Viaud-Delmon, I. (2016). *Human Computer Confluence. Transforming Human Experience Through Symbiotic Technologies.* Warsaw. De Gruyter Open.

Giger, J. C., Piçarra, N., Alves-Oliveira, P., Oliveira, R., & Arriaga, P. (2019). Humanization of robots: Is it really such a good idea? *Human Behavior and Emerging Technologies, 1*(2), 111-123.

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition, 125*(1), 125-130.

Leveringhaus, A. (2018). Developing robots: The need for an ethical framework. *European View, 17*(1), 37-43.

Marchetti, A., Manzi, F., Itakura, S., & Massaro, D. (2018). Theory of mind and humanoid robots from a lifespan perspective. *Zeitschrift für Psychologie, 226*(2), 98-109.

Marchetti, A., Di Dio, C., Manzi F., & Massaro D. (2020). Robotics in clinical and developmental psychology. In Riva G. (Ed.), *Reference Module in Neuroscience and Biobehavioral Psychology* (pp. 1-19), Elsevier.

Piçarra, N., & Giger, J.-C. (2018). Predicting intention to work with social robots at anticipation stage: Assessing the role of behavioral desire and anticipated emotions. *Computers in Human Behavior, 86,* 129-146.

Riva, G., & Wiederhold, B. K. (2021). Human-robot confluence: Toward a humane robotics. *Cyberpsychology, Behavior, and Social Networking, 24*(5), 291-293.

Sandini, G., & Sciutti, A. (2018). Humane Robots – from Robots with a humanoid body to robots with an anthropomorphic mind. *Journal of Human-Robot Interaction, 7*(1), Article 7.

Sciutti, A., Mara, M., Tagliasco, V., & Sandini, G. (2018). Humanizing human-robot interaction: On the importance of mutual understanding. *IEEE Technology and Society Magazine, 37*(1), 22-29.

Vignolo, A., Noceti, N., Rea, F., Sciutti, A., Odone, F., & Sandini, G. (2017). Detecting biological motion for human-robot interaction: A link between perception and action. [10.3389/frobt.2017.00014]. *Frontiers in Robotics and AI, 4*, 14.

SECTION 1

From Human-Robot Interaction to Human-Robot Experience

A Theoretical Framework

1. Towards Human-Robot Shared Experience

The Role of Social and Interpersonal Dimensions in Shaping Human-Robot Collaboration

A. Gaggioli, A. Chirico, D. Di Lernia, M.A. Maggioni, C. Malighetti, F. Manzi, A. Marchetti, D. Massaro, F. Rea, D. Rossignoli, G. Sandini, D. Villani, B.K. Wiederhold, G. Riva, A. Sciutti

ABSTRACT

Collaborative robotics is a fast-growing area with a vast range of applications, such as healthcare, small manufacturing, entertainment and sports. For robot collaborators to meet human needs and expectations in these domains, however, general social abilities are required. These are necessary for machines to handle real-life interactive contexts, over also extended periods of time. Collaboration is a complex relational process that entails mutual understanding and reciprocal adaptation and where social norms play a relevant role. This shift in focus from 'human-robot interaction' to 'human-robot shared experience' implies that the central focus of modeling should be on constructs such as co-adaptation, intersubjectivity, individual differences, and identity. In this contribution, we suggest that the emergence of a shared experiential space between humans and robots requires an evolution of current mainstream functional design approaches. More specifically, over and above functional robot capabilities, future architectural frameworks are needed that integrate the enabling dimensions of social cognition.

Introduction

Although interest in human-robot interaction and the development of socially competent machines is growing, this is often limited to very contextualized, short-term instances. Examples include the bartender robot, the guide robot in museums, the robot sales assistant or the robot receptionist – where the interaction is cursory and bound to the specific domain competencies.

The work contained in this chapter is an extension of the article originally published as Gaggioli, A., Chirico, A., Di Lernia, D., Maggioni, M., Manzi, F., Marchetti, A., Massaro, D., Rossignoli, D., Sandini, S., Villani, D., Riva, G., Sciutti, A. (2021). Machines like us and people like you: Towards human-robot shared experience. *Cyberpsychology, Behaviour, and Social Networking, 24*(5), 357-361.

A common belief about general purpose, socially competent robotics, is that their realization is limited by sensor performance or hardware/processing capabilities. Now, however, substantial advances have been attained in materials, actuators, sensors, and computational power. Indeed, this has brought about important improvements in the physical and computational abilities of current technologies. We have now robots which can do parkour and computational system able to solve the most complex logical challenges (e.g. winning at chess, Go or even Starcraft), by exploiting recent Artificial Intelligence solutions. However, despite all these advances, current social robotics is still far from the general abilities we expect a robot collaborator to possess. Effective collaboration between humans stems from 'growing together', i.e., from building a mutual understanding, that evolves over long periods through shared experiences.

Here we argue that the introduction of effective robot collaborators hinges upon the development of an intersubjective space between humans and machines. This is a requirement for shifting from considering robots just as sophisticated tools, or prostheses, to acknowledging robots as autonomous agents able to collaborate with us. Such a change in perspective is necessary if we want to share our responsibilities and duties with a robot, to delegate everyday tasks to it without continuous direct control. An activity as simple as going out to buy a carton of milk at a grocery store is a great challenge for an autonomous robot, if it is not able to model the complexity of that social context. This could include understanding the intentions of a passer-by's movements to properly navigate in a socially aware way, addressing the correct seller in the shop, and following more sophisticated social and pragmatic rules, e.g., giving precedence to an elderly or pregnant customer, or appropriately greeting familiar bystanders. That is why we wish to develop artificial entities that are capable of autonomy, mutual understanding, empathy, and ultimately relational skills. We believe that this will make possible future autonomous robots who can *collaborate with* us in everyday life, rather than simply executing our explicit commands or being designed to just work *close* with humans.

Why Are Shared Experiences Key to the Development of Collaborative Robots?

In recent times, we have witnessed an increase of activity to develop robots to be used outside the research labs and heavy manufacturing plants, suggesting the possibility to use robots in 'normal human environments' like homes, hospitals, and care centers. Indeed, the future

application of robots has been extended into social interactions with 'fragile' human beings (such as the elderly, children, and in general people in need of care with no specific experience with robots; Kang et al., 2020; Wood et al., 2021).

In parallel to robots entering human social spaces, humans have also begun to enter areas traditionally populated by robots alone, such as the manufacturing industry, where now so-called co-bots (an abbreviation of cooperative robots) are replacing classical independent robots, as human-robot co-working has proved to be more versatile and effective.

Despite this growing interest in robotic collaboration, creating real collaborative machines is challenging. Collaboration is a broad concept, which is used to describe a wide variety of behaviors where more than one agent works on a single task (Roschelle & Teasley, 1995). We would argue that collaboration is not just agents working side by side on complementary tasks, but also involves the establishment of mutual understanding and co-adaptation.

In robotics, co-adaptation is generally regarded as the adaptation to the skills of the user over time, to potentially trigger a corresponding adaptation in the human fellow operator (Nikolaidis et al., 2017). However, skills and actions are only a component of the relational processes involved when two humans collaborate. Similarly, human-robot interaction should embrace complex co-adaptation, where also the perceptual, affective, and cognitive dimensions dynamically change and somehow merge in a mutually transformative, shared experience (Tanevska et al., 2020). Following this reasoning, the bi-directionality of the process becomes central – as the unit of analysis should not be the individual, but the emerging system represented by the dyad or group (Sandini et al., 2018).

However, we do not consider co-adaptation as the only key dimension to developing collaborative robots. Insights into developmental science, philosophy of the mind, and behavioral economics point to further relevant dimensions in collaboration, which include intersubjectivity, individual differences, and identity.

Intersubjectivity

Studies in lifespan show that preferences and acceptability of robots in different contexts (Marchetti et al., 2020 a, b) are related to the like-me nature of social robots, as a function of the developmental level, concerning physical features and behaviors (Marchetti et al., 2018; Di Dio et al., 2020a,b,c; Manzi et al., 2020a,b; Manzi et al., 2021a,b). Based on this assumption, several studies have shown that robots' human resemblance

in terms of physical characteristics, as in humanoid robots, triggers both children and adults to anthropomorphize them. In a recent study, Manzi et al. (2020b) examined in 5-, 7- and 9-year-old children the attribution of mental states to two humanoid robots: NAO – more anthropomorphic – and Robovie – more mechanical (Phillips et al., 2018). The results showed that younger (5-year-old) children tended to anthropomorphize both robots, i.e., ascribing similar human psychological qualities to them. Older children, on the other hand, ascribed significantly greater human mental characteristics to the anthropomorphic robot than to the mechanical one, showing a greater sensitivity compared to younger children to the level of the robot's anthropomorphization. These results reveal an important developmental effect on the phenomenon of anthropomorphization of robotic agents. Sensitivity to the robot's human resemblance as a function of development also shapes the adult's attribution of mental qualities to robots. A recent study by Manzi et al. (2021b) analyzed the attribution of mental states by young adults to two humanoid robots, NAO and Pepper, which varied in their level of physical anthropomorphization. The results showed that a robot's anthropomorphism also affects adult attributions of psychological qualities to robots, as evidenced by the higher attribution of mental states to Pepper (more like an adult) compared to NAO. These results have important implications, not only for the understanding of human-robot interactions, but also the design of robots, which should consider the age of the target user.

Although, from childhood, we have a number of expectations of robots before we interact with them (as demonstrated by the studies on the attribution of mental qualities to the robots), the construction of intersubjectivity also requires real human-robot interactions.

The construction of intersubjectivity is an essential step for developing more human-like exchanges between humans and robots. As in humans, repeated interactions build a common history through the sharing of subsequent experiences, characterized also by errors, mismatches, and relational reparations (Marchetti et al., 2018; Di Dio et al., 2020b). This process, in turn, creates a relational memory and generates expectations around the relational experiences. For instance, we tend to help those who have helped us in the past and who could help us in the future. As such, reciprocity is one of the key elements sustaining the emergence and maintenance of cooperation between individuals and within social groups, both in adults and children (Zonca et al., 2021a,b). Understanding under which conditions robots are expected to comply to reciprocity or to similar social norms, in contrast to what happens in the case of computers, is currently an active subject of research (Zonca et al., 2021c,d).

In the novel *Machines Like Me* by Ian McEwan, the new generation an-

droid named Adam becomes part of Charlie's life. An intense relationship, full of the contradictions typical of human relationships, is established between the two main protagonists and Miranda, the third vertex in the love triangle: Adam profoundly influences Charlie's life. At the same time, Adam – set up by Charlie (and Miranda) with some initial parameters of personality and living the daily experience of human relationships – gradually changes his way of interacting with Charlie, developing new ideas about the world and 'him'self. The author depicts the phenomenon of co-adaptation between humans and robots: the two agents, one human and the other artificial, share different experiences, modifying their way of experiencing themselves and the world.

In envisaging a human-robot relationship, as imagined by Ian McEwan, it is essential that the first form of intersubjectivity (Stern, 2004) is established, to imagine a sharing of experiences between humans and machines. Among the psychological aspects that can facilitate the construction of relationships between humans, and even more so between humans and robots which are not yet familiar entities, is certainly trust. Coming back again to McEwan's novel, Charlie's trust in the android Adam undergoes abrupt changes in the story, especially when Adam falls in love with Charlie's partner. After this event it is difficult for Charlie to continue trusting Adam. The novel highlights that even trust between a robot and a human is influenced by dynamics much like those occurring between humans. Obviously, McEwan's novel depicts a future that is probably still far from the current scenario of complex relationships with robots. However, it is crucial to understand from the outset how trusting relationships are built when they include at least one robotic agent, especially if the robot becomes a companion or mentor to our child, or a health care assistant to one of our close relatives.

We suggest that the robotic agent represents a cultural and material artifact, which influences the individual's psychological development. Vygotsky's cultural-historical theory (1978) explained well how culture shapes the lines of development of our intelligence, through cultural and material artifacts. Social robots are not exempt from this type of influence, as they are themselves cultural and material artifacts and, therefore, can contribute to directing the psychological development of the individual in an original and, in many aspects, innovative way. In this perspective, the possible relationship between humans and robots and the emergence of intersubjectivity becomes a natural outcome.

Another important author in developmental psychology who helps shed light on the possible co-adaptation between humans and robots is Daniel Stern, one of the foremost experts in studying the ontogeny of intersubjectivity. Stern (2004) suggests that intersubjectivity is a need and, at the same time, a fundamentally human condition: our mind, by

its nature, is constantly seeking other people with whom to resonate and share experiences.

Two aspects emerge from Stern's hypothesis: the interpersonal dynamics that regulate intersubjectivity; and the motivational elements that drive human beings to enter a relationship with others. The first concerns how the robot would position itself in interpersonal terms with the human in this co-adaptation logic. Research has shown that provisional assimilation of the robot into a range of intersubjective dynamics is possible, as highlighted by a recent study by Manzi et al. (2020a). In this work, a robot that simulates salient social behaviors, such as eye-gaze, can trigger social expectations in humans, starting from the first months of life. Consequently, it generates an intersubjective space in infants that have not yet experienced the complexity of relational dynamics. A further eye-tracker study by Manzi et al. (Under Review) with 17-month-old toddlers revealed that, from the earliest months of life, they can anticipate the actions of a robotic and of a human partner, showing the ability to extend agency also to robots. Moreover, the study indicated that children activate more cognitive resources to anticipate the robot's actions than human ones, as demonstrated by the greater attention paid to the robot. This finding suggests that, from early childhood, humans can activate cognitive resources to understand the dynamic relationships of even an unfamiliar robotic partner.

Also in interaction with adults, robots can evoke responses similar to those observed in social exchanges among humans. This has been recently demonstrated by measuring the neural responses during the observation of a humanoid expressing a positive or negative attitude toward their partner through its movement (Di Cesare et al., 2020). By acting gently or rudely, the robot could evoke a similar activation in the dorsal-central insula of the observer as that measured when witnessing analogous human actions. The proper selection of human-inspired social behaviors enables a robot to provide information about an attitude or an affective state, a form of communication that has been named *vitality forms* by Daniel Stern (2010).

The second issue concerns the 'fundamental human condition' highlighted by Stern. Indeed, humans are born with a set of capacities that are modified through experience and learning. Two key questions arise in this context: what basic equipment should be implemented in the robot to establish an intersubjective space with humans? And: is it only necessary for the robot to simulate human skills or should the robot be equipped with some basic skills that can be developed autonomously through experience and learning? Again, research has shown that some typically human processes, such as trust, are fundamental for building and maintaining relationships with a robot, even in early childhood in short-term in-

teractions (Di Dio et al., 2020b). However, studies have shown that behavior simulation is only sufficient to elicit relational engagement in short-term interactions with robots (Aroyo et al., 2018), whereas it is desirable to equip the robot with skills that develop autonomously through long-term interactions with human partners and the world (Cross et al., 2019; Manzi et al., 2021a). Here we come back to Vygotsky, highlighted by Marchetti et al. (2018), who introduced the concept of the Zone of Proximal Development (1978), which represents an intersubjective space between two subjects of the relationship. In this zone, cognitive discrepancy of one subject can become a powerful motivator of co-adaptation. From this perspective, one of the most intriguing challenges for human-robot interactions is the possibility for robots to provide their human partners with stimuli, inputs, and interpretations that moderately exceed the partner's current capabilities, building bridges to more advanced forms of shared understanding and capabilities.

Dialogic Interaction

It is worth noting that most recent social robots can talk in a way similar to human beings. This feature has been the object of extensive research in the robotics domain (e.g., Broz et al., 2014) and is particularly relevant for collaboration, since the ability to communicate through words constitutes a distinctive feature of human interactions. Words (spoken or written) are the main means of communication among human beings and are crucial in complex interactions that need dialogue as a coordination tool, especially when the other's behavior does not correspond to one's own expectations. A long tradition in Philosophy, dating back to Socrates (469-399 BC), has stressed the ontological foundations of dialogic relationships. Martin Buber (1958) assumed that a full understanding of one's own identity is strictly dependent on dialogue with another presence.

Dumouchel and Damiano (2016) and Damiano and Dumouchel (2018) show that dialogue is the fundamental structure and basic pattern of how humans act and think: "The real anthropomorphism in social robotics derives from basic cognitive structures and in particular from our tendency to teleological thought and dialogue as the main form of interaction" (Dumouchel & Damiano, 2016; p. 110). This may explain why humans tend to treat artifacts (and humanoid artifacts) as interlocutors/partners. This is especially true for artifacts such as social robots, which should be able (in the framework of collaborative robots expressed before) to express and/or perceive emotions, communicate with high-level verbal dialogue, learn/recognize models of other agents, establish/maintain social relationships, use natural cues (gaze, gestures,

etc.), exhibit distinctive personality and character, and learn/develop social competencies (Fong et al., 2003).

Recent behavioral economics literature shows that the decision to trust a partner or to act in a trustworthy way – despite being incompatible with the rationality assumption of economic theory – is rather common across experimental subjects. Social robots are very useful in devising experiments able to explain the emergence of cooperation beyond the traditional approaches of self-regarding preferences (e.g., repeated games with the infinite horizon or with a finite uncertain horizon) or others regarding preferences (e.g., equity and/or fairness-based preferences). In most of this literature (see, among others: Collins et al., 2016; Zanatto et al., 2019; Oliveira et al., 2020), robots are used, instead of confederate human agents, to implement a multi-arm experiment in which subjects are exposed to several different types of partners. The use of robots assures the experimenter that no uncontrolled individual variation in the confederate agents' behavior could influence the results.

Such an approach allows one to also explore trust in the context of immersive interactive games and analyze the impact on trust of behaviors unique to robots, e.g., mechanical failures (Aroyo et al., 2021).

Maggioni and Rossignoli (2021), in contrast, have devised an experimental setting in which human subjects (university students) are matched to either a human or anthropomorphic partner and asked to perform a repeated prisoner's dilemma in which the randomly assigned treatment requires the Operator to react verbally to a social sub-optimum reached in the first stage, before the subject was asked about their willingness to play a second stage. Their preliminary results show that a verbal dialogic interaction with the robotic partner, who verbally reacts to the actions of the human player in a simple repeated co-operation game, reduces the otherwise negative bias that human subjects show toward robot partners when compared with human partners. Twenty-five years ago, in their seminal paper, Farrell and Rabin (1995, p. 108) stated that "people don't usually take the destructively agnostic attitude of 'I won't presume that the words mean what they have always meant'. Rather, people take the usual or literal meaning seriously". The preliminary findings in Maggioni & Rossignoli (2021) show that, somehow, what is true for human-human interactions, holds also for human-robot interactions.

Individual Factors

In the process of co-adaptive development of the human-robot relationship, it is important to consider that different outcomes are influenced by multiple variables. At the individual level, personality has been iden-

tified as a core factor for understanding the nature and quality of this relationship (Robert et al., 2020). Personality refers to "those characteristics of the person that account for consistent patterns of feelings, thinking, and behaving" (Pervin, 2005, p. 6) and it explains the way people respond and interact with others in social settings.

In brief, researchers have found a positive impact of human personality – especially personality traits according to the Big Five taxonomy (McCrae & Costa, 2008) – on various human-robot interaction outcomes, including distance and approaching direction, perceptions, and attitudes towards the robot, emotion towards the robot, anthropomorphism, and trust (Syrdal et al., 2007; Reich-Stiebert & Eyssel, 2015; Conti et al., 2017; Robert, 2018). Specifically, extraverted people tend to be more sociable and to talk more with robots (Ivaldi et al., 2017), are more willing to interact with robots (Syrdal et al., 2007; Salam et al., 2016) and report higher levels of confidence in interacting with robots compared to introverts (Haring, Matsumoto & Watanabe, 2013). Individuals with high levels of openness to experience are more oriented to accept assistive robotic technologies (Conti et al., 2017). People characterized by emotional instability report high levels of stress and maladaptive coping when interacting with autonomous agents (Szalma & Taylor, 2011) and prefer to be physically distanced from the robot (Takayama & Pantofaru, 2009).

Furthermore, research has also considered the contribution of different personality characteristics of robots on the human-robot relationship. Specifically, several studies have highlighted that extroverted and socially intelligent robots were more often preferred in terms of acceptableness, trustworthiness, and enjoyableness (De Ruyter et al., 2005; Looije et al., 2010; Tay, Jung & Park, 2014). Why is a robot's extraversion the most studied trait? Because it is a core trait involved in social interactions and can be easily manipulated. Typically, extraversion is associated with greater eye contact with humans (Andrist et al., 2015) and increased vocal dialogue and body movements in response to human behavior (Windhouwer, 2012; Aly & Tapus, 2013; Celiktutan & Gunes 2015). Thus, eye contact, vocal dialogue and gestures can be manipulated in robots to interpret this personality trait.

Finally, the third area of investigation is constituted by the analysis of human-robot personality similarities and differences (match and mismatch). In this regard, several studies found a positive effect of personality match on the quality of interaction, in terms of enjoyment and engagement, social attraction, credibility, trust, and compliance (Joosse et al., 2013; Niculesco et al., 2013; Andriella et al., 2020). However, it is important to recognize that similarity does not always predict the best interaction. Other important variables, such as the nature of the task, the

role of the robot, human expectations or the appearance of the robot can play a crucial role on the quality of the interaction. Furthermore, the time people interacted with the robot in the past research was mainly of short duration. For the future, it would be important to analyze how repeated interactions influence the perception of personality over time (Paetzel-Prüsmann et al., 2021).

Research has shown that, besides personality traits, other individual differences affect both interactions and interpersonal relationships with robots, as well as expectations of technical and interactional features being implemented in robots. Different studies analyzed the effect on human-robot interactions of people's negative attitudes towards robots; they revealed that repeated interactions can reduce the levels of anxiety experienced towards robots (Nomura et al., 2008) and that the attitudes are a moderator factor for the effects of social presence (Schellen & Wykowska, 2019). Specifically, the greater the human negative attitudes towards robots, the less social influence robots exert in interactive games on their human partners (Hinz et al., 2019). In other words, people's attitudes towards robots can positively or negatively influence co-adaptation between humans and robots, shaping a specific intersubjective space that is fundamental for sharing experiences and for relational dynamics.

Attitudes towards robots are also an important predictor of people's expectations of relational skills being implemented in robots. A recent study showed that the expectations of young adults can be placed along a continuum of robot humanization and that negative attitudes towards robots can reveal the type of expectations in terms of humanization (Manzi et al., 2021c). The results showed that more positive attitudes towards robots are not necessarily associated with a greater desire to implement robot relational skills. These findings stress that individual differences are crucial for capturing how people differ qualitatively in their mental models for advanced robots and for understanding the different co-adaptation forms between humans and robots. Thus, although human-robot relationships develop dynamically after only short interactions (Edwards et al., 2019), it is important to consider some individual factors that come into play and contribute coherently and adaptively to the co-construction of the human-robot relationship.

Identity

Although a key component of co-adaptation is the possibility to change, following the context, events, and needs of the partner, this poses a risk. If an agent constantly changes, it is impossible to define its *identity*.

In other words, the human will no longer be able to build an understanding of the partner and anticipate it, practically destroying the possibility for an interaction (Sciutti et al., 2018). Indeed, identity is a key component of a relational architecture. On the one hand, it cannot be conceived as a static, preprogrammed feature, otherwise, it would hinder co-adaptation; on the other hand, providing unconstrained adaptability to the robot would prevent the evolution of stable behavioral features, that are necessary building blocks to develop and sustain the relationship, by eliciting positive familiarity (Finkel et al., 2015). Since identity in humans derives from a complex interplay of genetic, environmental, social, and cultural factors, determining how this should be intended for robots is a task that goes beyond the scope of this short reflection. However, this is a problem that will need to be tackled if we are to build human-robot shared experiences. The intrinsic dialogic nature of the human being has long been debated in philosophy. Martin Buber in his book *I and Thou* (Buber, 1970) argues that a full understanding of one's own identity (the I) is strictly dependent on dialogue with another presence (the Thou). The heteronomous revelation of a singular presence calls the subject into an open-ended relationship. At the core of this model of existence is the notion of 'encounter' as a revelation of 'presence' (*Gegenwart*). In contrast to 'object' (*Gegenstand*), the presence revealed by an encounter occupies the space 'in-between' the subject and another. This 'in-between' space is defined as 'mutual' (*Gegenseitig*). This stance prompts a deeper exploration of the concept of 'encounter' in the context of human-robot shared experiences.

Conclusion

Collaborative robotics is a field of research that presents important technical and technological challenges. However, it also represents a unique opportunity to explore closely the relational and social dynamics that can be established among humans and robots.

We suggest that the following dimensions should be taken into closer consideration: the co-adaptation between agents, the emergence of intersubjectivity, and the role of individual differences and identity. All together these could participate in the achievement of human-robot shared experiences. The latter are necessary because, while currently human partners grow, change, and learn from interaction with each other, the robot is left to its capabilities, without the skill to adapt and form shared experiences with partners. This happens, because most current cognitive architectures are designed to allow a robot to intelligently operate in the environment (Kotseruba & Tsotsos, 2020) with

limited or non-existent modeling of the components of social and affective cognition. This gap becomes evident when robots are faced with real-life contexts and activities, which often span extended temporal periods. From a methodological viewpoint, aiming for human-robot shared experiences would imply, on the one hand, changing the existing 'solipsistic' approach to architecture design; on the other, developing more inclusive evaluation frameworks, which extend the focus from the functional aspects of human-robot interaction to its social experiential dimension.

To conclude, in our view, the relational space between humans and robots won't be manually crafted by human engineers. Rather it might be enabled by humans and then emerge through interaction. In this way, robots may participate in this development as co-protagonists. Accordingly, although our approach to developing robots endowed with social capabilities necessarily stems from a human-centered epistemology, we contemplate the possibility that the collaboration of humans and robots may lead to the emergence of a novel, 'porous' epistemology – one contaminated by the perspective of the robots themselves.

References

Aly, A., & Tapus, A. (2013). A model for synthesizing a combined verbal and non-verbal behavior based on personality traits in human-robot interaction. In *8th ACM/IEEE international conference on human-robot interaction (HRI)*, Tokyo, Japan (pp. 325-332).

Andriella, A., Siqueira, H., Fu, D., Magg, S., Barros, P., Wermter, S., Torras, C., & Alenya, G. (2020). Do I have a personality? Endowing care robots with context-dependent personality traits. *International Journal of Social Robotics*, 1-22.

Andrist, S., Mutlu, B., & Tapus, A. (2015). Look like me: Matching robot personality via gaze to increase motivation. In *33rd annual ACM conference on human factors in computing systems, Seoul Republic of Korea* (pp. 3603-3612).

Aroyo, A. M., Rea, F., Sandini, G., & Sciutti, A. (2018). Trust and social engineering in human robot interaction: Will a robot make you disclose sensitive information, conform to its recommendations or gamble? *IEEE Robotics and Automation Letters, 3*(4), 3701-3708.

Aroyo, A. M., Pasquali, D., Kothig, A., Rea, F., Sandini, G., & Sciutti, A. (2021). Expectations vs. reality: Unreliability and transparency in a treasure hunt game with iCub. *IEEE Robotics and Automation Letters, 6*(3), 5681-5688.

Broz, F., Nehaniv, C. L., Belpaeme, T., Bisio, A., Dautenhahn, K., Fadiga, L., Ferrauto, T., Fischer, K., Förster, F., Gigliotta, O., Griffiths, S., Lehmann, H., Lohan, K. S., Lyon, C., Marocco, D., Massera, G., Metta, G., Mohan, V., Morse, A., Nolfi, S., & Cangelosi, A. (2014). The ITALK project: A developmental robotics approach to

the study of individual, social, and linguistic learning. *Topics in cognitive science, 6*(3), 534-544.

Buber, M. (1970). *I and Thou* (W. Kaufmann, Trans.). Charles Scribner's Sons (Original work published 1923).

Celiktutan, O., & Gunes, H. (2015). Computational analysis of human-robot interactions through first-person vision: Personality and interaction experience. In *(RO-MAN), 24th IEEE International Symposium on Robot and Human Interactive Communication*, Kobe, Japan (pp. 815-820).

Collins, M. G., Juvina, I., & Gluck, K. A. (2016). Cognitive model of trust dynamics predicts human behavior within and between two games of strategic interaction with computerized confederate agents. *Frontiers in psychology, 7*, 49.

Conti, D., Commodari, E., & Buono, S. (2017). Personality factors and acceptability of socially assistive robotics in teachers with and without specialized training for children with disability. *Life Span and Disability, 20*(2), 251-272.

Cross, E. S., Hortensius, R., & Wykowska, A. (2019). From social brains to social robots: Applying neurocognitive insights to human-robot interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences, 374*(1771), 20180024.

Damiano, L., & Dumouchel, P. (2018). Anthropomorphism in human-robot co-evolution. *Frontiers in Psychology, 9*, 468.

De Ruyter, B., Saini, P., Markopoulos, P., & Van Breemen, A. (2005). Assessing the effects of building social intelligence in a robotic interface for the home. *Interacting with computers, 17*(5), 522-541.

Di Cesare, G., Vannucci, F., Rea, F., Sciutti, A., & Sandini, G. (2020). How attitudes generated by humanoid robots shape human brain activity. *Scientific Reports, 10*(1), 1-12.

Di Dio, C., Manzi, F., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., & Marchetti, A. (2020a). It does not matter who you are: Fairness in pre-schoolers interacting with human and robotic partners. *International Journal of Social Robotics, 12*(5), 1045-1059.

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., & Marchetti, A. (2020b). Shall I trust you? From child-robot interaction to trusting relationships. *Frontiers in psychology, 11*, 469.

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., & Marchetti, A. (2020c). Come i bambini pensano alla mente del robot: il ruolo dell'attaccamento e della Teoria della Mente nell'attribuzione di stati mentali ad un agente robotico [How children think about the robot's mind. The role of attachment and Theory of Mind in the attribution of mental states to a robotic agent]. *Sistemi Intelligenti, 1*(20), 41-56.

Dumouchel, P., & Damiano, L. (2016). *Vivre avec les robots. Essai sur l'empathie artificielle.* [*Living with Robots, Essay about Artificial Empathy*]. Le Seuil.

Edwards, A., Edwards, C., Westerman, D., & Spence, P. R. (2019). Initial expectations, interactions, and beyond with social robots. *Computers in Human Behavior, 90*, 308-314.

Farrell, J. (1995). Talk is cheap. *The American Economic Review, 85*(2), 186-190.

Finkel, E. J., Norton, M. I., Reis, H. T., Ariely, D., Caprariello, P. A., Eastwick, P. W., Frost, J. H., & Maniaci, M. R. (2015). When does familiarity promote versus undermine interpersonal attraction? A proposed integrative model from erstwhile adversaries. *Perspectives on Psychological Science, 10*(1), 3-19.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems, 42*(3-4), 143-166.

Haring, K. S., Matsumoto, Y., & Watanabe, K. (2014). Perception and trust towards a lifelike android robot in Japan. In H. Kim, S. I. Ao, M. Amouzegar (Eds.), *Transactions on Engineering Technologies* (pp. 485-497). Springer.

Hinz, N.A., Ciardo, F., & Wykowska, A. (2019). Individual differences in attitude toward robots predict behavior in human-robot interaction. In M. Salichs et al. (Eds.), *Lecture Notes in Computer Science: vol 11876. Social Robotics. ICSR 2019* (pp. 64-73). Springer.

Ivaldi, S., Lefort, S., Peters, J., Chetouani, M., Provasi, J., & Zibetti, E. (2017). Towards engagement models that consider individual factors in HRI: On the relation of extroversion and negative attitude towards robots to gaze and speech during a human-robot assembly task: Experiments with the iCub humanoid. *International Journal of Social Robotics, 9*(1), 63-86.

Joosse, M., Lohse, M., Pérez, J. G., & Evers, V. (2013). What you do is who you are: The role of task context in perceived social robot personality. In *2013 IEEE International Conference on Robotics and Automation* (pp. 2134-2139). IEEE.

Kang, H. S., Makimoto, K., Konno, R., & Koh, I. S. (2020). Review of outcome measures in PARO robot intervention studies for dementia care. *Geriatric Nursing, 41*(3), 207-214.

Kotseruba, I., & Tsotsos, J. K. (2020). 40 years of cognitive architectures: core cognitive abilities and practical applications. *Artificial Intelligence Review, 53*(1), 17-94.

Looije, R., Neerincx, M. A., & Cnossen, F. (2010). Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors. *International Journal of Human-Computer Studies, 68*(6), 386-397.

Maggioni, M. A., & Rossignoli, D. (Forthcoming). Look who's talking: Dialogue and trust in human-robot interactions.

Maggioni, M.A., & Rossignoli, D. (2021). If it looks like a human and speaks like a human... dialogue and cooperation in human-robot interactions, *ArXiv*.

Manzi, F., Di Dio, C., Di Lernia, D., Rossignoli, D., Maggioni, M., Massaro, D., Marchetti, A., & Riva, G. (2021a). Can you activate me? From robots to human brain. *Frontiers in Robotics and AI, 8*, 633514.

Manzi, F., Massaro, D., Di Lernia, D., Maggioni, M., Riva, G., & Marchetti, A. (2021b). Robots are not all the same: young adults' expectations, attitudes, and mental attribution to two humanoid social robots. *Cyberpsychology, Behavior, and Social Networking, 24*(5), 307-314.

Manzi, F., Sorgente, A., Massaro, D., Villani, D., Di Lernia, D., Malighetti, C., Gaggioli, A., Rossignoli, D., Sandini, G., Sciutti, A., Rea, F., Maggioni, M.A., Marchetti, A., & Riva, G. (2021c). Emerging adults' expectations about the next generation of robots: Exploring robotic needs through a latent profile analysis. *Cyberpsychology, Behavior, and Social Networking*, *24*(5), 315-323.

Manzi, F., Ishikawa, M., Di Dio, C., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., & Marchetti, A. (2020a). The understanding of congruent and incongruent referential gaze in 17-month-old infants: An eye-tracking study comparing human and robot. *Scientific Reports*, *10*, 11918.

Manzi, F., Peretti, G., Di Dio, C., Cangelosi, A., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., & Marchetti, A. (2020b). A robot is not worth another: Exploring children's mental state attribution to different humanoid robots. *Frontiers in Psychology*, *10*, 2011.

Manzi, F., Ishikawa, M., Di Dio, C., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., & Marchetti, A. (Under Review). Infants' prediction of robot's goal-directed action. *International Journal of Social Robotics*.

Marchetti, A., Di Dio, C., Massaro, D., & Manzi, F. (2020a). The psychosocial fuzziness of fear in the COVID-19 era and the role of robots. *Frontiers in Psychology, 10*, 2245.

Marchetti, A., Di Dio, C., Manzi, F., & Massaro, D. (2020b). Robotics in clinical and developmental psychology. *Reference Module in Neuroscience and Biobehavioral Psychology*.

Marchetti, A., Manzi, F., Itakura, S., & Massaro, D. (2018). Theory of mind and humanoid robots from a lifespan perspective. *Zeitschrift für Psychologie*, *226*(2), 98-109.

McCrae, R. R., & Costa, P. T. Jr. (2008). The five-factor theory of personality. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of Personality: Theory and Research* (pp. 159-181). The Guilford Press.

Niculescu, A., van Dijk, B., Nijholt, A., Li, H., & See, S. L. (2013). Making social robots more attractive: the effects of voice pitch, humor and empathy. *International journal of social robotics, 5*(2), 171-191.

Nikolaidis, S., Hsu, D., & Srinivasa, S. (2017). Human-robot mutual adaptation in collaborative tasks: Models and experiments. *The International Journal of Robotics Research, 36*(5-7), 618-634.

Nomura, T., Kanda, T., Suzuki, T., & Kato, K. (2008). Prediction of human behavior in human-robot interaction using psychological scales for anxiety and negative attitudes toward robots. *IEEE transactions on robotics, 24*(2), 442-451.

Oliveira, R., Arriaga, P., Santos, F. P., Mascarenhas, S., & Paiva, A. (2020). Towards prosocial design: A scoping review of the use of robots and virtual agents to trigger prosocial behaviour. *Computers in Human Behavior, 114*, 106547.

Paetzel-Prüsmann, M., Perugia, G., & Castellano, G. (2021). The influence of robot personality on the development of uncanny feelings. *Computers in Human Behavior*, 120, 106756.

Pervin, L. A., & John, O. P. (2001). *Personality: Theory and Research (9th edition).* John Wiley & Sons.

Pfeiffer, T., Rutte, C., Killingback, T., Taborsky, M., & Bonhoeffer, S. (2005). Evolution of cooperation by generalized reciprocity. *Proceedings of the Royal Society B: Biological Sciences, 272*(1568), 1115-1120.

Phillips, E., Zhao, X., Ullman, D., & Malle, B. F. (2018). What is human-like? Decomposing robots' human-like appearance using the anthropomorphic robot (abot) database. *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*, 105-113.

Reich-Stiebert, N., & Eyssel, F. (2015). Learning with educational companion robots? Toward attitudes on education robots, predictors of attitudes, and application potentials for education robots. *International Journal of Social Robotics, 7*(5), 875-888.

Robert, L. (2018). Personality in the human robot interaction literature: A review and brief critique. *Proceedings of the 24th Americas Conference on Information Systems*, 16-18.

Robert, L., Alahmad, R., Esterwood, C., Kim, S., You, S., & Zhang, Q. (2020). A review of personality in human-robot interactions. *Foundations and Trends® in Information Systems, 4*(2), 107-212.

Roschelle, J., & Teasley, S. D., (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (ed.), *Computer Supported Collaborative Learning: vol 128. NATO ASI Series* (pp. 66-97). Springer.

Salam, H., Celiktutan, O., Hupont, I., Gunes, H., & Chetouani, M. (2016). Fully automatic analysis of engagement and its relationship to personality in human-robot interactions. *IEEE Access, 5*, 705-721.

Sandini, G., Mohan, V., Sciutti, A., & Morasso, P. (2018). Social cognition for human-robot symbiosis – challenges and building blocks. *Frontiers in neurorobotics, 12*, 34.

Schellen, E., & Wykowska, A. (2019). Intentional mindset toward robots – open questions and methodological challenges. *Frontiers in Robotics and AI, 5*, 139.

Sciutti, A., Mara, M., Tagliasco, V., & Sandini, G. (2018). Humanizing human-robot interaction: On the importance of mutual understanding. *IEEE Technology and Society Magazine, 37*(1), 22-29.

Stern, D.M. (2004). *The Present Moment in Psychotherapy and Everyday Life.* W. & W. Norton & Company.

Stern, D. N. (2010). *Forms of Vitality Exploring Dynamic Experience in Psychology, Arts, Psychotherapy, and Development.* Oxford University Press.

Syrdal, D. S., Koay, K. L., Walters, M. L., & Dautenhahn, K. (2007). A personalized robot companion? The role of individual differences on spatial preferences in HRI scenarios. In *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 1143-1148). IEEE.

Szalma, J. L., & Taylor, G. S. (2011). Individual differences in response to automation: The five-factor model of personality. *Journal of Experimental Psychology: Applied, 17*(2), 71.

Takayama, L., & Pantofaru, C. (2009). Influences on proxemic behaviors in human-robot interaction. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 5495-5502). IEEE.

Tanevska, A., Rea, F., Sandini, G., Cañamero, L., & Sciutti, A. (2020). A socially adaptable framework for human-robot interaction. *Frontiers in Robotics and AI, 7*, 121.

Tay, B., Jung, Y., & Park, T. (2014). When stereotypes meet robots: The double-edge sword of robot gender and personality in human-robot interaction. *Computers in Human Behavior, 38*, 75-84.

Vygotsky, L. S. (1978). *Mind in society: The Development of Higher Psychological Processes*. Harvard University Press.

Windhouwer, D. (2012). The effects of the task context on the perceived personality of a Nao robot. In *Proceedings of the 16th Twente student conference on IT*. Enschede.

Wood, L. J., Zaraki, A., Robins, B., & Dautenhahn, K. (2021). Developing Kaspar: A humanoid robot for children with autism. *International Journal of Social Robotics, 13*(3), 491-508.

Zanatto, D., Patacchiola, M., Goslin, J., & Cangelosi, A. (2019). Investigating cooperation with robotic peers. *PloS one, 14*(11), e0225028.

Zonca, J., Folsø, A., & Sciutti, A. (2021a). Dynamic modulation of social influence by indirect reciprocity. *Scientific reports, 11*(1), 1-14.

Zonca, J., Folsø, A., & Sciutti, A. (2021b). I'm not a little kid anymore! Reciprocal social influence in child-adult interaction. *Royal Society Open Science, 8*(8), 202124.

Zonca, J., Folsø, A., & Sciutti, A. (2021c). The role of reciprocity in human-robot social influence. *iScience*, 103424.

Zonca, J., Folsø, A., & Sciutti, A. (2021d). Trust is not all about performance: Trust biases in interaction with humans, robots and computers. *arXiv preprint arXiv:2106.*14888.

2. Relationship Development with Humanoid Social Robots

Applying Interpersonal Theories to Human-Robot Interaction

J. Fox, A. Gambino

ABSTRACT

Humanoid social robots (HSRs) are human-made technologies that can take physical or digital form, resemble people in form or behavior to some degree, and are designed to interact with people. A common assumption is that social robots can and should mimic humans, such that human-robot interaction (HRI) closely resembles human-human (i.e., interpersonal) interaction. Research is often framed from the assumption that rules and theories that apply to interpersonal interaction should apply to HRI (e.g., the computers are social actors framework). Here, we challenge these assumptions and consider more deeply the relevance and applicability of our knowledge about personal relationships to relationships with social robots. First, we describe the typical characteristics of HSRs available to consumers currently, elaborating characteristics relevant to understanding social interactions with robots such as form anthropomorphism and behavioral anthropomorphism. We also consider common social affordances of modern HSRs (persistence, personalization, responsiveness, contingency, and conversational control) and how these align with human capacities and expectations. Next, we present predominant interpersonal theories whose primary claims are foundational to our understanding of human relationship development (social exchange theories, including resource theory, interdependence theory, equity theory, and social penetration theory). We consider whether interpersonal theories are viable frameworks for studying HRI and human-robot relationships given their theoretical assumptions and claims. We conclude by providing suggestions for researchers and designers, including alternatives to equating human-robot relationships to human-human relationships.

This chapter was originally published as 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. This research was not funded.

Introduction

Defining social robots as 'humanoid' implies that these machines are intended to be perceived similarly to people. Both designers and researchers of human-robot interaction (HRI) often rely on human-human interactions as models or standards for how to build and study robots (Bickmore & Picard, 2005; de Melo & Gratch, 2015; Krämer et al., 2012; Nass & Brave, 2005). Although advances in engineering and artificial intelligence have made robots more human like, their communicative and social capacities are still relatively primitive, inhibiting human-robot relationship development. A common belief is that robots will eventually become sophisticated enough that they will be indistinguishable from humans, but this assumption begets two questions. First, is existing theorizing about interpersonal, human-human relationships applicable to studying human-robot relationships? And second, given what we know, should HRI designers' goal be to mimic them?

To address these issues, first we explicate humanoid social robots (HSRs), their features, and the common social affordances they currently offer. Then, we explore the extent to which these robots meet the assumptions of theories of interpersonal relationship development. We consider whether interpersonal theories are viable frameworks for studying HRI and human-robot relationships now and in the future. Finally, we provide some suggestions for researchers and designers.

Humanoid Social Robots

We define *humanoid social robots* as human-made technologies that can take physical or digital form, resemble people in form or behavior to some degree, and are designed to communicate with people (Breazeal, 2002; Hegel et al., 2009; Zhao, 2006). Examples of HSRs include conversational agents (e.g., chatbots or voice assistants such as Siri and Alexa; Liu & Sandar, 2018), embodied conversational agents (e.g., virtual coaches or health care providers; Cassel et al., 2000), consumer robots that specialize in education and home care (e.g., Zora robot), and robots that are designed primarily to interact with humans (e.g., Pepper). This conceptualization excludes robots that lack resemblance to humans or do not interact socially and semiautonomously with humans, such as industrial robots, robotic home appliances (e.g., Roomba), self-driving cars, and telepresence robots (e.g., Beam, Double Robotics).

The anthropomorphic characteristics of HSRs may prompt human users to treat HSRs in human-like ways (Di Dio et al., 2020; Kahn et al., 2013). One of the most popular theoretical frameworks adopted in the

study of human-computer interaction (HCI) and HRI is the computers are social actors perspective (CASA), derived from the media equation (Nass & Moon, 2000). According to these perspectives, technology has outpaced biological evolution: human brains have not evolved to identify and distinguish mediated simulations. Instead, humans react mindlessly and naturally, responding to a media representation in the same way they would respond to its natural counterpart (Nass & Moon, 2000).

CASA argues that computers can demonstrate the potential for social interaction through anthropomorphic appearance cues (e.g., having a human-like face or form) or behavior (e.g., using language or bipedal locomotion). Human users respond naturally and mindlessly to these cues, treating the computer like another social being. Consequently, CASA claims that any rules or findings about human-human interactions should carry over to human-computer interactions if the computer demonstrates social cues (Nass & Moon, 2000). Some studies have tested CASA's claims with HSRs and found support. For example, an HSR with gendered facial cues can lead people to apply gender stereotypes (Eyssel & Hegel, 2012). In addition, if an HSR is put on the same team as a human, the human will like it more than an HSR from a different team.

The CASA paradigm has been used to justify hypotheses suggesting HRIs are comparable with interpersonal interactions and relationships (Bickmore & Picard, 2005; Edwards et al., 2019; Krämer et al., 2012); however, results from empirical studies do not consistently support CASA's predictions (Gambino et al., 2014). Media technologies have evolved considerably since the bulk of the original research and theorizing (including CASA's predecessor, the media equation; Reeves & Nass, 1996) emerged in the 1990s, presenting two challenges. First, early research within the paradigm was focused on interactions with simple computer interfaces or singular aspects of interfaces (e.g., voice). Extrapolating CASA's thesis to more complex, dynamic HSRs may not be appropriate. A second argument is that since the 1990s, people have considerably more experience with computers and robots. Thus, they have developed more specified social scripts for human-computer and human-robot interactions and are not applying human-human interaction scripts as CASA suggests (Gambino et al., 2014). Here, we suggest a third reason that findings and theories about human-human interactions do not necessarily apply to current HCI or HRI: modern social technologies such as HSRs are simply not sophisticated enough to fulfill the roles or perform the complex tasks that human social interactants do naturally.

Although robots will certainly become more flexible, sophisticated, and intelligent in the future, the types of HSRs that the average human consumer is likely to encounter in the coming years remain limited in

their capacities due to complexity, cost, and technological constraints. Social scientists seeking to understand and explain HRI must consider the current state of HSRs and adopt a practical perspective of the foreseeable future rather than relying on assumptions that technological advances in artificial intelligence and robotics will be so swift as to obviate the need for theorizing in the intervening years or decades. For this reason, we outline the typical characteristics and affordances of current, prevalent forms of HSRs rather than the most cutting edge robotic technologies that very few people have experienced or are likely to experience in the near future.

Characteristics and social affordances of modern HSRs

Although modern HSRs are defined by some ability to make decisions and take actions on their own, they are not fully autonomous (Beer et al., 2014; Bigman et al., 2019). Modern HSRs require a human user to initiate actions (e.g., through pressing buttons, writing a script, or launching a program) or interact, supervise, or intervene in the process (i.e., the human in the loop). Humans are also required to handle maintenance, such as recharging or cleaning, and manage any obstacles or technological problems the robot encounters.

Anthropomorphism and social affordances indicate potential to communicate with a user (Fox & McEwan, 2017), which is a definitive function of HSRs (Rodríguez-Hidalgo, 2020). Modern HSRs vary in their anthropomorphic characteristics, or the extent to which they resemble and are perceived as human (Epley et al., 2007; Leite et al., 2013; Ruijten et al., 2019; van Straten et al., 2020). *Form anthropomorphism* entails sensory cues that make a robot seem human-like. For example, HSRs may have a human-like voice or appearance. Even with high levels of form anthropomorphism, modern HSRs are unlikely to be mistaken for a human due to low levels of *behavioral anthropomorphism*, or the extent to which an HSR's actions resemble a human (e.g., gestures, spoken messages, nonverbal expressions) (Nowak & Fox, 2018).

Limitations in their social affordances and technological capacities hamper modern HSRs' ability to communicate in a human-like manner (Strohkorb et al., 2016; van Straten et al., 2020; Zhao, 2006). One major issue is that modern HSRs lack the ability to attend to, recall, and apply relevant information from previous interactions with a human user, which diminishes social perceptions (Aylett et al., 2013; Strohkorb et al., 2016). Robot memory lacks *persistence* and sufficiently refined *searchability* for sensible ongoing social interactions: most do not maintain a memory of previous interactions, and if they do, retrieval is constrained

to a few task-relevant queries. Modern HSRs cannot make sense of interactional history in the same way that humans do, and they are limited in their ability to apply such knowledge to novel social situations (Strohkorb et al., 2016).

Because modern HSRs are limited in the tasks they are programmed to perform, *interactivity* can be difficult to navigate. The robot is constrained to a small set of possible responses, limiting *responsiveness* and *contingency*, which can violate users' expectations and diminish feelings of closeness and trust (van Straten et al., 2020). Because HSRs are designed to cater to their human user and limited in their interactional abilities, they have minimal *conversational control* (Reeves & Nass, 1996). They lack the autonomy to change topics or tasks, or to interrupt or terminate interactions with human users. HSRs are created to satisfy the human user's needs, and the human does not need deviations or defiance.

Without a persistent memory and the ability to execute contingent actions, the robot is limited in its ability for *personalization*, or tailoring an interaction to a specific individual (Gambino et al., 2020). In personal relationships, humans shape their messages based on their previous knowledge and interactions with a target, which enhances feelings of closeness (Altman & Taylor, 1973). Similar personalization is expected and desired from HSRs (van Straten et al., 2020; Dautenhahn, 2004). Most HSRs cannot identify nor distinguish different users, however; it treats all users agnostic of individual variations or previous interactions. Even for robots that can recall some parameters for a specific user, this knowledge does not help personalize or tailor the message in real time based on the target's verbal and nonverbal cues (Strohkorb et al., 2016).

Collectively, these current limitations suggest that modern HSRs lack many of the fundamental social capabilities of humans. Even if HSRs can engage in some forms of social interaction, these limitations have implications for how interactions transpire over time and, importantly, the viability of developing relationships with humans.

Considerations for human-robot relationship research

Despite the long-standing shortcomings of HSRs (Hegel et al., 2009; Strohkorb et al., 2016; van Straten et al., 2020), a considerable amount of social scientific research, such as that from the CASA perspective (Nass & Moon, 2000), is designed, hypothesized, and conducted assuming HSRs are perceived similarly to humans. Another issue is that the majority of HRI research involves a single session of interaction and often with a technology that is novel to the user (Breazeal, 2002; Leite et

al., 2013). If users have no experience with a particular HSR, or with HSRs in general, they may be more likely to apply human-human scripts in their initial interactions due to what they perceive as the robot's social affordances. In this way, one-shot studies may suggest that, yes, claims about interpersonal relationships can carry over.

Importantly, however, a relationship is not a one-shot experience. As Hinde (1979) clarified, "A relationship involves a series of interactions between two individuals known to each other". Relationships are characterized by familiarity established through multiple engagements, which indicates studies with a single session are not well-suited for determining an HSR's potential for a human-like relationship (Breazeal, 2002).

Moreover, perceptions of social affordances change over time as a user becomes more familiar with a technology (Gambino et al., 2020), which has been noted in several longitudinal studies with social robots (Leite et al., 2013; van Straten et al., 2020). Often, these lead to expectancy violations when social robots cannot maintain human standards for communication and becoming acquainted (Leite et al., 2013). Here, we explore the extent to which modern HSRs are capable of developing relationships similar to human-human relationships and whether predominant theories of relational development can be applied to studying human-robot relationships.

Applicability of Interpersonal Theories of Relationship Development

The predominant paradigm for understanding interpersonal relationships is based on the concept of social exchange (Roloff, 1981). The fundamental assumptions of social exchange are that people need resources to survive, other people can provide resources, and sharing and trading resources is a fundamental aspect of relationships (Roloff, 1981). Although this perspective has been proposed as a valid foundation for understanding human-robot relationships (Krämer et al., 2012), a closer examination of various theories indicates that the nature of modern HSRs may challenge their assumptions and claims.

Resources

According to the resource theory of social exchange, the resources people exchange in relationships range in how tangible or abstract they are and what function they serve (Foa & Foa, 1974). For example, money or goods are tangible economic resources, whereas love, status, and information are more abstract and social. Tangible resources are trans-

ferred from one person to another; the giver must relinquish a resource such as money or goods, and ownership shifts from the giver to the recipient. Intangible resources, in contrast, are shared between two people (Roloff, 1981). Resources also vary on *particularity*, or how much a resource's value is contingent on who is providing it (Foa & Foa, 1974). For example, money spends the same whether it is received from a bank teller or a spouse, whereas love is particular and presumably more valuable coming from one's spouse rather than a bank teller.

HRI CHALLENGE. HSRs have some resources to provide and perhaps exchange with humans, particularly services and information. A robot itself, however, is a tangible good that is owned by someone, and as such does not have ownership over resources such as money or other goods and cannot transfer them to a human recipient. As HSRs are subservient to humans, they have little to offer humans in terms of status. Although it is possible for a human to form an attachment (Kahn et al., 2013), and perhaps love a robot, this is a unidirectional offering rather than a shared resource. Therefore, most social exchange resources outlined by resource theory may not be pertinent to evaluating human-robot relationships. In addition, a lack of persistent, personalized memory indicates that resources are not particularized from the robot's perspective, and it is likely the human would not perceive a robot's resources as particularized either. Thus, as relational partners, robots would be perceived as interchangeable and relationships with them impersonal rather than special.

Costs, benefits, and equity

Social exchange theories also posit that humans are fundamentally self-interested and evaluate the costs and benefits they incur in a relationship (Roloff, 1981; Thibaut & Kelley, 1959). Within a relationship, individuals become interdependent through their exchange of resources, which may be more or less symmetrical (Thibaut & Kelley, 1959). As such, relationships may be evaluated on whether these exchanges are relatively balanced, or if one partner is incurring more costs or receiving more benefits. These evaluations may be based on perceptions of equity, or whether an individual is receiving benefits or output proportional to the amount of input or costs they are incurring, particularly compared with the social norm (Walster et al., 1978). According to interdependence theory, one way to assess the costs and benefits of one's relationship is to compare it with other relationships, or to compare the relationship with the current partner to alternatives (Kelley & Thibaut,

1978; Thibaut & Kelley, 1959). If partners feel underbenefited, they are likely to experience dissatisfaction and seek to restore equity in the relationship. Over time, underbenefited partners may become dissatisfied with the relationship and terminate it, particularly if there are desirable alternatives.

HRI CHALLENGE. A major challenge to the applicability of social exchange theories is considering the nature of costs and benefits to robots. HSRs lack the motivation and desire that characterize human needs; they do not experience rewards, punishments, benefits, and costs in the ways that humans do. Furthermore, HSRs are servile to human controllers; there is a permanent inequity as they are designed to maximize benefits for humans with minimal, if any, consideration of the costs they might incur. They do not make autonomous evaluations of equity with their human controllers. They do not experience dissatisfaction, make comparisons, nor have the ability to act based on these assessments. They cannot leave their human controllers. Knowing this, humans do not have to consider their robot partner's costs or benefits, only the costs and benefits to themselves. They can make unilateral decisions without concern for the robot's wishes or well-being based on one self-serving principle: maximize my benefits and never mind the robot.

Self-disclosure

Social penetration theory (SPT; Altman & Taylor, 1973) suggests that relationships develop through reciprocal self-disclosure. Individuals consider the costs and benefits of ongoing disclosure and determine whether they want to intensify the relationship.

Altman and Taylor (1979) used an onion metaphor to explain how individuals maintain many layers self rooted in their experiences, in which the outside layer is the publicly observable self and private information is stored in deeper layers that must be uncovered. In a developing relationship, individuals peel back these layers through reciprocal self-disclosure, proceeding through stages characterized by expanding breadth and growing depth. Breadth is characterized as the range of topics or categories that comprise the self, such as social identities, interests, and experiences. Depth involves the beliefs and values that are central to the self (Altman & Taylor, 1979).

According to SPT, the earliest stage of a relationship, orientation, is characterized by small talk and governed by social norms of appropriateness (Altman & Taylor, 1979). In the exploratory affective stage, slightly deeper self-disclosure occurs across a broader range of topics. In the

affective exchange stage, feelings of intimacy escalate as partners reveal deeper facets of the self, including values, goals, or fears. The stable exchange stage is characterized by mutual understanding, and partners are comfortable disclosing deep private matters.

HRI CHALLENGE. Given modern HSRs are constrained in their tasks and abilities, they do not have much breadth. HSRs also do not have a unique cluster of beliefs, values, and self-image that characterize depth. Although HSRs may share information, it is not based on personal experience or self-image; thus, exchanges with HSRs arguably do not qualify as self-disclosure, and they could not engage in the reciprocal self-disclosure required in a developing relationship.

It should be noted that a human partner may make false attributions about the social potential of HSRs, particularly in short-term interactions. As SPT notes, social norms guide early interactions and disclosures are shallow (Altman & Taylor, 1979). Humans are more likely to follow scripts of socially acceptable behavior that may be easier for HSRs to mimic, and researchers may then observe effects similar to what would be expected of a human-human interaction. Over time, however, humans would recognize an HSR's lack of personalized persistent memory, which would be necessary to build a relationship.

In summary, common HSRs are not human like enough at this time to meet the fundamental assumptions and claims of key interpersonal theories, and it is unclear when, or if, they ever will be. These challenges the popular mindset of working from the assumption that human-human findings will apply to HSRs and applying our understanding of interpersonal interactions to HRI and human-robot relationships.

Discussion

Collectively, these issues indicate that researchers must carefully consider their theoretical options for studying contemporary human-robot relationships. One is to propose and test modifications of or extensions to existing interpersonal theories to accommodate HSRs. By examining intervening variables such as perceived agency, behavioral anthropomorphism, and perceived social affordances, researchers may be able to broaden the utility and scope of existing interpersonal theories. Alternatively, given modern HSRs violate interpersonal theories' fundamental assumptions of humanity, scholars could consider human-human interaction an inherent boundary condition of these theories and shift to developing and testing new models founded in human-robot relationships that may or may not explain human-human relationships. Either way,

to ensure the validity of human-robot relationship research, it is crucial for scholars to conduct studies with multiple interactions over time and to account for experience and existing familiarity (Bigman et al., 2019; Dautenhahn, 2007).

Regardless of their current application, existing interpersonal theories could serve as a form of Turing test in the future study of human-robot relationships. If HSRs are eventually designed to meet theoretical assumptions, these theories can then be tested to see if they are upheld in the human-robot relationship development. The fit of human-human relationship theories would indicate that HSRs have achieved greater similarity to humans.

Yet, should being perceived as humans be the goal of HSR design? Generally HSRs are being created to complement or augment human capacity, performing tasks to serve humans (Fong et al., 2002). By design, such HSRs will never have the same power or autonomy as the humans who program, own, and control them. Theoretically, it seems problematic to adopt an agnostic perspective about this inherent difference between humans and robots. Ethically, it seems problematic to humanize robots and encourage human-like relationships with objects that are under the control of, exist at the whim of, and are designed to satisfy their human owner. One concern is that if users develop social scripts with humanoid robots, these scripts could be applied to human-human interactions and lead to the objectification, dehumanization, or mistreatment of other people (de Graaf, 2016; Kahn et al., 2013; Scheutz, 2012).

Indeed, humans can be quite terrible, which also calls into question whether human-human interactions are always optimal models. Social exchange theories claim that humans are inherently self-interested; must robots be? Humans rely on oft-detrimental stereotypes and implicit biases to evaluate other people; could robots overcome this deficit? Lacking human-like characteristics could also be beneficial in particular contexts. For example, if a person is disclosing sensitive or stigmatizing information, they may fear being harshly judged, stereotyped, or rejected by a human due to existing social norms. Studies have shown that people disclose more to an HSR than a human, perhaps because an HSR may seem less judgmental or people may feel more anonymous interacting with a robot (Kang & Gratch, 2010; Lucas et al., 2017; Pickard et al., 2016).

Given their fundamental differences, one possibility is designing and studying social robots through different lenses than human-human relationships. One suggested model has been human-pet, or more specifically human-dog, relationships (Dautenhahn, 2004; de Graaf & Allouch, 2017). Another proposed approach is conceptualizing robots

more broadly as human companions (Krämer et al., 2011). Although these models may be viable in some contexts, social robots are distinct and warrant their own theorizing, particularly given the growing variation in the roles they play.

If we accept robots as unique social beings, we do not need to refer to them as 'humanoid.' Indeed, HRI designers should explore novel ways social robots could interact, relate, and bond beyond human abilities and norms. Designers should consider ways robots may be uniquely suited to maximize positive social outcomes (Riva et al., 2012) or minimize negative ones. Such advancements may expand and illuminate not only human-robot relationships but also human-human relationships.

In conclusion, theories for understanding human-human relationships are likely unsuitable for examining modern human-robot relationships, given the current HSRs' shortcomings as social actors. These approaches may in fact be restrictive, as social robots may be able to compensate for human shortcomings or exceed human capacity in some ways. Going forward, researchers must continually reevaluate the emerging features and social affordances of robots to understand human-robot relationships now and in the future.

References

Altman I., & Taylor, D. A. (1973). *Social Penetration: The Development of Interpersonal Relationships*. Holt, Rinehart & Winston.

Aylett, R., Kriegel, M., Wallace, I., Segura, E. M., Mecurio, J., Nylander, S., & Vargas, P. (2013). Do I remember you? Memory and identity in multiple embodiments. In *2013 IEEE RO-MAN* (pp. 143-148). IEEE.

Beer, J. M., Fisk, A. D., & Rogers, W. A. (2014). Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of human-robot interaction, 3*(2), 74.

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI), 12*(2), 293-327.

Bigman, Y. E., Waytz, A., Alterovitz, R., & Gray, K. (2019). Holding robots responsible: The elements of machine morality. *Trends in cognitive sciences, 23*(5), 365-368.

Breazeal, C. (2002). *Designing Sociable Robots*. MIT Press.

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. F. (2000). *Embodied Conversational Agents*. MIT Press.

Dautenhahn, K. (2004). Robots we like to live with?! A developmental perspective on a personalized, life-long robot companion. In *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759)* (pp. 17-22). IEEE.

Dautenhahn, K. (2007). Methodology & themes of human-robot interaction: A growing research field. *International Journal of Advanced Robotic Systems, 4*(1), 15.

de Graaf, M. M. (2016). An ethical evaluation of human-robot relationships. *International journal of social robotics, 8*(4), 589-598.

de Graaf, M. M. A., & Allouch, S. B. (2017). The influence of prior expectations of a robot's lifelikeness on users' intentions to treat a zoomorphic robot as a companion. *International Journal of Social Robotics, 9*(1), 17-32.

de Melo, C. M., & Gratch, J. (2015). Beyond believability: quantifying the differences between real and virtual humans. In *International Conference on Intelligent Virtual Agents* (pp. 109-118). Springer.

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., & Marchetti, A. (2020). Shall I trust you? From child-robot interaction to trusting relationships. *Frontiers in psychology, 11*, 469.

Edwards, A., Edwards, C., Westerman, D., & Spence, P. R. (2019). Initial expectations, interactions, and beyond with social robots. *Computers in Human Behavior, 90*, 308-314.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological review, 114*(4), 864.

Eyssel, F., & Hegel, F. (2012). (S)he's got the look: Gender stereotyping of robots. *Journal of Applied Social Psychology, 42*(9), 2213-2230.

Foa, U. G., & Foa, E. B. (1974). *Societal Structures of the Mind*. Charles C. Thomas.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. A. (2002). Survey of socially interactive robots: concepts, design, and applications. *Robotics and Autonomous Systems, 42*,142-166.

Fox, J., & McEwan, B. (2017). Distinguishing technologies for social interaction: The perceived social affordances of communication channels scale. *Communication Monographs, 84*(3), 298-318.

Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication, 1*(1), 5.

Hegel, F., Muhl, C., Wrede, B., Hielscher-Fastabend, M., & Sagerer, G. (2009). Understanding social robots. In *2009 Second International Conferences on Advances in Computer-Human Interactions* (pp. 169-174). IEEE.

Hinde, R. A. (1979). *Towards Understanding Relationships*. Academic Press.

Kahn Jr, P. H., Gary, H. E., & Shen, S. (2013). Children's social relationships with current and near-future robots. *Child Development Perspectives, 7*(1), 32-37.

Kang, S. H., & Gratch, J. (2010). Virtual humans elicit socially anxious interactants' verbal self-disclosure. *Computer Animation and Virtual Worlds, 21*(3-4), 473-482.

Kelley, H. H., & Thibaut, J. W. (1978). *Interpersonal Relations: A Theory of Interdependence*. Wiley.

Krämer, N. C., Eimler, S., Von Der Pütten, A., & Payr, S. (2011). Theory of compan-

ions: What can theoretical models contribute to applications and understanding of human-robot interaction?. *Applied Artificial Intelligence, 25*(6), 474-502.

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. In M. D. Zacarias & J. V. Oliveira (Eds.), *Human-computer Interaction: The Agency Perspective* (pp. 215-240). Springer.

Leite, I., Martinho, C., & Paiva, A. (2013). Social robots for long-term interaction: A survey. *International Journal of Social Robotics, 5*(2), 291-308.

Liu, B., & Sundar, S. S. (2018). Should machines express sympathy and empathy? Experiments with a health advice chatbot. *Cyberpsychology, Behavior, and Social Networking, 21*(10), 625-636.

Lucas, G. M., Rizzo, A., Gratch, J., Scherer, S., Stratou, G., Boberg, J., & Morency, L. P. (2017). Reporting mental health symptoms: Breaking down barriers to care with virtual human interviewers. *Frontiers in Robotics and AI, 4*, 51.

Nass, C. I., & Brave, S. (2005). *Wired for Speech: How Voice Activates and Advances the Human-computer Relationship*. MIT Press.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of social issues, 56*(1), 81-103.

Nowak, K. L., & Fox, J. (2018). Avatars and computer-mediated communication: A review of the definitions, uses, and effects of digital representations. *Review of Communication Research, 6*, 30-53.

Pickard, M. D., Roster, C. A., & Chen, Y. (2016). Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars and under what conditions?. *Computers in Human Behavior, 65*, 23-30.

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

Riva, G., Baños, R. M., Botella, C., Wiederhold, B. K., & Gaggioli, A. (2012). Positive technology: using interactive technologies to promote positive functioning. *Cyberpsychology, Behavior, and Social Networking, 15*(2), 69-77.

Rodríguez-Hidalgo, C. (2020). Me and my robot smiled at one another: The process of socially enacted communicative affordance in human-machine communication. *Human-Machine Communication, 1*(1), 4.

Roloff, M. E. (1981). *Interpersonal Communication: The Social Exchange Approach*. Sage.

Ruijten, P. A., Haans, A., Ham, J., & Midden, C. J. (2019). Perceived human-likeness of social robots: testing the Rasch model as a method for measuring anthropomorphism. *International Journal of Social Robotics, 11*(3), 477-494.

Scheutz M. (2012). The inherent dangers of unidirectional emotional bonds between humans and social robots. In P. Lin, J. Abney & J. A. Bekey (Eds.), *Robot Ethics: The Ethical and Social Implications of Robotics* (pp. 205-222). MIT Press.

Strohkorb, S., Huang, C. M., Ramachandran, A., & Scassellati, B. (2016). Establish-

ing sustained, supportive human-robot relationships: Building blocks and open challenges. In *2016 AAAI Spring Symposium Series* (pp. 179-182). IEEE.

Thibaut, J. W., & Kelley, H. H. (1959). *The Social Psychology of Groups*. Wiley.

van Straten, C. L., Peter, J., & Kühne, R. (2020). Child-robot relationship formation: A narrative review of empirical research. *International Journal of Social Robotics, 12*(2), 325-344.

Walster, E., Walster, G. W., & Berscheid, E. (1978). *Equity*. Allyn & Bacon.

Zhao, S. (2006). Humanoid social robots as a medium of communication. *New Media & Society, 8*(3), 401-419.

3. A Conceptual Characterization of Autonomy in the Philosophy of Robotics

C. De Florio, D. Chiffi, F. Fossa

ABSTRACT

The concept of autonomy is crucial for the theoretical characterization of robots and, more in general, complex technological artifacts. The aim of this paper is to provide a conceptual and logical framework in which it is possible to define two concepts of autonomy: autonomy of performance and autonomy of process. The analysis is carried out exploiting the logical resources of the counterfactual semantics – developed by Lewis' and Stalnaker's seminal works – and branching structures of the possible courses of actions. It allows to differentiate the autonomy of a robotic arm from the autonomy of a highly complex machine-like rovers for the explorations of the planets. The clarification of the concept of autonomy is, moreover, an essential precondition for the investigation concerning the ethics of artificial agents.

Introduction

Developing ethical reflections has become an urgent and unavoidable need in the philosophy of robotics. The increasingly widespread use of robots and, more generally, artificial agents in many practical contexts raises important ethical questions. It would, however, be misleading to think that the philosophy of robotics coincides with the ethics of robotics. On the contrary, theoretical reflection on robotics is necessary to provide solid foundations for applied ethical analysis. By 'theoretical reflection' we refer to areas of investigation whose main aim is to elaborate rigorous philosophical accounts of the most characteristic features of robotic artifacts. From this perspective, one of the more pressing and controversial topics is, without doubt, that of robot autonomy.

Shedding light on robot autonomy is also crucial for the ethics of robotics, to the point that the notion is mainly discussed in inquiries on robot or machine ethics. Indeed, in order to understand how ethics and robotics are intertwined it is necessary to account for the constitutive features of robot agency, which in turn cannot be done without a thorough analysis of the kind of autonomy robots are commonly believed to

enjoy. Indeed, autonomy is the very reason why (some) robots are usually classified as artificial agents, i.e., special artefacts to which we are naturally prone to assign some, if not all, of the characteristics of more traditional agents like human beings. Robotic functioning itself – unlike the functioning of many other artifacts and tools – is commonly framed as an agent's action (Maes, 1991; Franklin & Grasser, 1995). It is precisely this peculiarity of artificial agents that opens a new ethical dimension which comprises the critical evaluation of both their use and their own actions.

The aim of this essay is to inquire into the notion of autonomy as it is usually ascribed to robots, in order to clarify some of its most characteristic features. Such a clarification is a necessary step towards developing a thorough account of what artificial agents are and why they pose such relevant ethical issues.

The methodology we adopted to explore the notion of autonomy was inspired by formal analysis and by a comparison between different explanatory hypotheses using a mainly abductive method (Magnani, 2001; Woods, 2013; Williamson, 2016; Chiffi & Pietarinen, 2020). However, our theoretical sketch will be aligned with a series of intuitions stemming from our current interaction with artificial agents, tools and robots. In order to take such intuitions into due consideration, we will introduce and discuss two case studies, one for each aspect of functional autonomy that we seek to clarify. Even though such intuitions may have a limited justificatory or inferential role, they can still guide us in identifying problems that are generally dealt with in a logical-philosophical context. It is well-known that one of the beneficial aspects of the adopted methodology is to include and harmonize certain intuitions and linguistic practices; yet, it is also worth noting that sometimes our *prima facie* beliefs will have to be renegotiated in view of new empirical results or of a different theoretical analysis. Such a dynamic recalls what John Rawls, in another context, defined as 'reflective equilibrium'. Moreover, the adopted methodology has the advantage of respecting the constraints of rationality and objectivity, while allowing for a fruitful interdisciplinary dialogue between technology, human sciences, and philosophy.

The paper is organized as follows. In section 2 we present some *preliminary considerations* about the notion of autonomy as it applies to robots, and the methodology we adopted to try and clarify some of its most significant features. In section 3 we present the *formal frameworks* we used to analyze this form of autonomy. Section 4 is dedicated to autonomy of *performance*, while in section 5 we focus on *autonomy of process*. Section 6 concludes the paper. We believe that these two models of robot autonomy represent a good starting point to its study as one of the key concepts in the philosophy of robotics.

Preliminary Considerations

As a first step, it is important to clarify that the term 'autonomy' has taken on a very specific meaning in the field of (philosophical) robotics. As such, it must not be confused with the concept commonly used to characterize human agency. A useful starting point for familiarizing oneself with its unique meaning is provided by Franklin and Grasser (1995), who describe a software agent's autonomy as the ability to pursue goals embedded in its design while sensing some aspects of the environment and exercising some control over its own action.

As Sullins (2006, p. 25) puts it, roboticists use the term to indicate robots that are "capable of making at least some of the major decisions about their actions using their own programming". Similarly – albeit within the multifaceted framework of the Levels of Abstraction methodology – Floridi (2011, p. 357) describes autonomy as the agent's ability "to change state without direct response to interaction", meaning that "it can perform internal transitions to change its state". However, as Verdicchio (2017) notes, this notion of autonomy as control over some decisions or internal states must always be understood as applied to robotic functioning, which already presents some relevant features[1] – the key one being that robotic functioning is a way to automate tasks. Consequently, robotic autonomy can only be properly understood through reference to the goal(s) or purpose(s) the system is designed to pursue (Amigoni & Schiaffonati, 2005). Indeed, "computer and robotic systems were originally introduced to perform specific tasks, so it should not come as a surprise that goals play a fundamental role in this context" (Verdicchio, 2017, p. 186). Autonomous robots – and, more generally, artificial agents – are "purpose-built artefacts" (Bryson, 2010; see also Bryson & Kime, 2011, p. 1), entities that are built precisely to carry out given functions that their builders value. As such, artificial agents are part of wider socio-technical contexts, within which they "have meaning and significance only in relation to human beings" (Johnson, 2011, p. 168).

In light of this, two points must then be underlined. First, the meaning commonly associated with the term 'autonomy' in robotics must be understood against this task-related background: autonomy is a term that describes how some robots execute given functions and accomplish

[1] In this respect, it is useful to go back to the distinction between *automaticity* and *autonomy* as discussed, for instance, by Steels (1995). It would be interesting to discuss if this distinction still holds when applied to machine learning algorithms or if it needs further specifications in order to properly account for the difference(s) between biological and artificial self-regulating behavior.

given goals. Secondly, this notion of autonomy is specific to robotic behavior, so it is necessary to thoroughly differentiate it from the notion of autonomy as ascribed to human agency. As Alan Winfield (2012, p. 24) explains:

> When roboticists talk about autonomous robots, they normally mean robots that decide what to do next entirely without human intervention or control. We need to be careful here because they are not talking about true autonomy, in the sense that you or I would regard ourselves as self-determining individuals, but what I would call 'control autonomy'. By control autonomy I mean that the robot can undertake its task, or mission, without human intervention, but that mission is still programmed or commanded by a human.

In light of this, robotic autonomy can also be defined as *functional autonomy*: the execution of given functions or the accomplishment of given goals marks the context or dimension to which robot autonomy belongs. In robotics, then, autonomy ascriptions are always embedded in a task-related, functional dimension[2]. Within this dimension, autonomy indicates the ability of an agent to execute one or more given tasks or functions by itself, without requiring constant human supervision or intervention (Tamburrini, 2020, p. 55)[3].

Such a characteristic renders robotic autonomy a unique phenomenon indeed. In a sense, it might be argued that this form of goal-oriented or task-related autonomy is actually merely the occurrence of a full-fledged form of heteronomy – i.e., not a defective form of autonomous agency (as in the case of human beings), but rather an unprecedented phenomenon in its own right. Of course, human agents can also execute given tasks or commands, so that task-related autonomy in reference to human agency can be discussed as well. However, we execute given tasks *as fully autonomous agents* – the type of autonomy human agents are commonly believed to enjoy. Order interpretation, evaluation, application, and acknowledgement are part of the task and are legitimately expected by those who give orders or assign tasks to humans. Only artificial agents, on the contrary, execute given tasks *as functional autonomous agents*, thus being fully heteronomous in their operations (Fossa, 2020).

Whether there is continuity or discontinuity between functional and full autonomy is a hotly debated issue in the field of superintelligence and general AI, which falls outside the scope of the present inquiry. Our purpose in what follows is to focus on the characteristic features of functional autonomy as a phenomenon that is proper to artificial agen-

[2] For a critical discussion of this 'instrumentalist' framework, see Gunkel (2012).

[3] For recent discussions on autonomy and artificial agency, see Wheeler (2020) and Müller (2021).

cy and, thus, represents a necessary step both in its theoretical and ethical study. This research, therefore, explores comparisons between humans and artificial agents not to investigate their continuity or discontinuity, but rather to shed light on the main features of functional autonomy understood as one of the key concepts in the philosophy of robotics.

A Formal Framework

Before focusing on a discussion of the various case studies and building the various models of autonomy, let us establish a logical framework that can simplify the characterization of the concepts[4].

Below are the ingredients of the framework:

i) A class of agents **A** whose elements are the following: a_1, a_2, a_3,...

ii) A class of tasks **T** whose elements are the following: t_1, t_2, t_3,…

Tasks may have any level of complexity (from tightening a screw, to adjusting the temperature, scoring a goal, etc.), and are linguistically defined by structures such as the infinitive or the '-ing' form. Each task may be easily associated with a corresponding proposition. Regulating the temperature, for example, can be translated as 'the fact that the temperature is regulated'.

iii) A class of factors **K**, whose elements are indicated as k_1, k_2, k_3,…

As we will see, *k*-factors are partial descriptions of the scenarios in which agents act. They consist of finite sequences of propositions, and can therefore be combined with Boolean connectives: k_1 & k_3, say, indicates the conjunction of the factors k_1 and k_3. Similarly, $\sim k$ indicates that *k*-factors do not hold.

iv) A ternary predicate of autonomy A(*a*, *t*, *k*), whose intended meaning is 'the agent *a* is autonomous in doing *t* with respect to factors *k*'.

In the next section we will observe how, starting from this basic predicate of autonomy, we can define some derivative predicates of autonomy that are able to characterize the intuitions present in the two case studies.

v) A binary agentive predicate *a*:*t* whose meaning is: agent *a* performs task *t*.

---

[4] For an overview on the logic of action, see Segerberg, Meyer and Kracht (2020); see also Belnap, Perloff and Xu (2001).

a:*t* is a factive relationship: if *a*:*t*, then *t* holds. On the other hand, it will be useful to introduce a *dispositional* version of this predicate, which we will refer to as ◊*a*:*t*, meaning that the agent *a* is able to perform task *t*, even if in the current circumstances the agent does not perform the action.

vi) A class **P** of procedures consisting of the elements p_1, p_2, p_3,...

Given the procedures, we can enrich our agentive relationship by forming a compound predicate *a*,*p*:*t*, which we can read as: agent *a* performs task *t* through procedure *p*.

At this point we have all the ingredients for a logical characterization of two key specifications of functional autonomy, i.e., autonomy of performance and autonomy of process.

Autonomy of Performance

As already mentioned, our interactions with artifacts such as instruments, tools, and robots presume, oftentimes implicitly and unconsciously, a series of assumptions about the structural properties of such entities. In our case, we will start by considering some common attributions (or non-attributions) of autonomy about such instrumental entities. Analyzing case studies that revolve around these (non-)attributions will help us refine the concept of functional autonomy by breaking it down into two different sub-types: *autonomy of performance* and *autonomy of process*.

Let us begin with autonomy of performance. With this label we wish to refer to the capacity of an artificial agent to carry out a given function without requiring the intervention or supervision of other agents. In this respect, this form of autonomy characterizes the execution of tasks carried out by agents that have (or are able to acquire) what it takes to do so, without needing external support. How could this intuitive idea be formulated in formal language according to the framework we introduced in the previous section?

In order to answer this question, let us focus on the following scenario:

Case Study 1. Sheila is a worker whose job is to tighten screws into a metal cover inside an assembly line. She uses a screwdriver and takes a considerably long time to tighten the screws into the bearings. A couple of years later, the company where Sheila works decides to provide its employees with more advanced tools, thus Sheila is given an electric screw-

driver: processing times decrease significantly as well as the amount of physical effort that Sheila needs to employ. For a couple of years now, the new owner has decided to use a robotic arm that fastens the screws into the metal cover. Sheila has been relocated to another department of the company.

We have all attempted at least once to tighten a screw. Without the appropriate tool, this task becomes very difficult (if not impossible) and good results are hard to achieve. Up to a point, we depend on our tools to execute this task, and our autonomy is somewhat constrained by such dependence. Nonetheless, when humans use tools, it is only natural to ascribe autonomy to them. Different tools, however, facilitate the task of tightening screws in different ways. Consider the regular screwdriver, the electric screwdriver, and the robotic arm in Sheila's story. Intuitively, only in the latter case would we say that the task has been fully automated and, as a consequence, that the robotic arm is autonomous in executing the task. Similarly, only in the latter case can the task be delegated in its entirety to the tool. In all other cases, human intervention and supervision is still necessary – even though the electric screwdriver requires less effort on the human part, since it automates the task only partially.

Although task automation comes in degrees, functional autonomy is intuitively only ascribed to tools to which it is possible to fully delegate a task or group of tasks. Accordingly, in our scenario, autonomy is intuitively ascribed to entities that are able to tighten screws by themselves: either Sheila or the robotic arm. While regular and electric screwdrivers still require Sheila's support for the task to be carried out, robotic arms carry out the task without relying on anybody's assistance. Of course, one can easily imagine situations in which human assistance might be needed even in the case of the robotic arm: malfunctioning, reprogramming, and so on[5]. Software and hardware issues aside, and all else being equal, it however makes sense to say that a robotic arm is autonomous in reference to screw tightening if it is able to execute the task without requiring other agents' constant intervention. Indeed, there is no job left for Sheila. She can be relocated or fired.

The whole idea behind this very intuitive and basic notion of auton-

[5] Russel and Norvig (2010, p. 39) state that learning should be counted among the fundamental components of autonomy, so that the execution of a given task can be said to be autonomous only if the agent is able to learn and adapt its initial knowledge to new situations and challenges. However, learning does not seem to be a prerequisite in the case of our robotic arm. It follows that functional autonomy of performance can be significantly ascribed to artificial agents even if there is no learning involved. As discussed in the following section, learning seems to play a more important role in the case of autonomy of process.

omy is that an agent is autonomous with respect to a task if such agent is able to complete the task even in the *absence* of other relevant factors. Autonomy is therefore understood as *agentive independence*. Agentive independence is the necessary condition to *autonomy of performance*, i.e., the ability to carry out a given task without having to rely on relevant external factors. This form of autonomy can be meaningfully attributed to our robotic arm and, more generally, to artificial agents. Thus, it represents a first characterization of functional autonomy.

Let us turn now to our formal framework. Our notion of autonomy of performance implies three actors: agent (a), task (t) and factor (k) with respect to which the agent may (or may not) be autonomous in completing the task. For now, let us consider other agents as factors. Suppose we wish to establish whether the robotic arm (a) is autonomous in screw tightening (t) with respect to Mark the technician, who is supervising its operations (k) after Sheila was relocated. How could this be done?

One way to illustrate more thoroughly the idea of autonomy of performance is by using counterfactual analysis (cf. Stalnaker & Thomason, 1970; Lewis, 1973). A counterfactual is a conditional in which the antecedent is not true in the current situation. For the evaluation of counterfactuals, we must refer to scenarios different from the current one. As a general rule, counterfactuals of the type 'if it were p then it would be q' (in symbols, $p \mathrel{\Box\!\!\rightarrow} q$), are evaluated by considering all the scenarios most similar to the present-day world in which p is true, and assessing if in such scenarios q is also true.

How does the counterfactual framework work in the example of our robotic arm? The robot is autonomous in tightening a screw with respect to Mark if and only if in all scenarios (more similar to the present-day world) in which Mark is not there, the robot is nonetheless able to tighten the screw.

(1) $\mathrm{A}(\mathrm{Robot}, t, k_m)$ iff $\sim k_{\mathrm{m}} \mathrel{\Box\!\!\rightarrow} \Diamond \mathrm{Robot}{:}t$

Conversely, the robot is *not* autonomous in tightening a screw with respect to Mark if and only if in all the scenarios (more similar to the present-day world) in which Mark is not present, the robot is unable to tighten the screw.

(2) $\sim\mathrm{A}(\mathrm{Robot}, t, k_m)$ iff does not hold that $\sim k_m \mathrel{\Box\!\!\rightarrow} \Diamond \mathrm{Robot}{:}t$

Counterfactual analysis can give us an account of the idea of agentive independence that is closer to our intuitions. Of course, it becomes now extremely important to clarify which factors are *relevant* to autonomy of performance ascriptions. Properly speaking, in fact, no agent

is completely autonomous: there are always material conditions to any form of autonomous behavior. In other words, there are always going to be *k* factors in the scenarios in which the agent performs an action. Think, for instance, of environmental conditions: if there were no oxygen, Sheila and Mark would probably not be able to do their job; if there were no electricity, the robotic arm would probably not be able to do its job in the long run. In a word, factors that are relevant to autonomy of performance ascriptions must be distinguished from irrelevant ones.

It is not easy to delineate a demarcation principle between relevant and irrelevant factors. As a first step, we can differentiate at least between *resources*, *tools*, and *other agents*. The relations between agents and these three factors in reference to a given task and for what concerns autonomy are evidently different.

Resources represent in most cases necessary conditions for the exercise of autonomy, but their unavailability would not lead to denying autonomy of performance to an agent. For example, we would not say that Sheila is not autonomous in reference to screw tightening were she unable to feed or breathe. Similarly, it would be at least odd to state that our robotic arm is not autonomous in screw tightening since it needs electricity to function. Resources can be described as factors but their unavailability does not seem to be all that relevant for the ascription of autonomy of performance.

Similarly, tools and related skills in their use are surely conditions to the execution of tasks but their availability is mostly presupposed in autonomy ascriptions in the first place. In fact, they are supposed to be available to an agent who *might have what it takes* to execute a task autonomously. This holds for both Sheila and the robotic arm. It is almost obvious that Sheila must be able to get a screwdriver and must know how to use it in order to be even considered as possibly autonomous in screw tightening. These are basic conditions that make her a good candidate in being autonomous in the execution of the task. In the case of the robotic arm, likewise, it would be nonsensical to consider it as a possible autonomous agent were the software or the CPU disconnected from the robotic body. The whole system is to be considered, even though without the effectors the computing units would not be able to carry out the task. So, it seems that tools and skills as factors are also not that relevant to our purposes.

In order to stick to our original intuition, we must therefore proceed with a restriction to certain subclasses of *k* factors: we are ready to recognize that the robotic arm is autonomous in screw tightening even if it requires electricity, for example. Also, we are ready to acknowledge autonomy to Sheila even if she needs screwdrivers and special skills to car-

ry out her task. What about situations where the other agents' intervention is required?

The need to rely on other agents' actions to carry out a task seems to be extremely relevant in autonomy of performance ascriptions. Again, this seems to hold in the case of task-related autonomy both for Sheila and the robotic arm. Should Sheila or the robotic arm need external support from other agents, then it would sound incorrect to ascribe autonomy in screw tightening to them. In this case, the agents' *dependence* on external factors appears to be a much more significant hurdle in the ascription of autonomy.

It is not hard to adjust our definition according to what has emerged. Let us introduce a factor k_A, which indicates the presence of other agents in the assessment scenario. Let us also refer to autonomy of performance as autonomy$_1$. We will therefore say that

(3) An agent a is autonomous$_1$ with respect to task t if and only if in all scenarios most similar to the present-day world in which *there are no other agents*, a is capable of performing task t:

$A_1(a, t)$ iff $\sim k_A \Box \rightarrow \Diamond a{:}t$

When this is the case and a robotic agent is involved, the robot can be said to be autonomous with reference to its task. The term indicates exclusively the ability of the technology to execute its specified task(s) without requiring the intervention of other agents. In this respect, no further ability is presupposed: we expect the robotic tool to behave as programmed – whatever program it implements – and its performance is evaluated according to these expectations. Suffice to mention here that expectations and evaluations are both significantly different when we move from functional autonomy to full autonomy scenarios, even if we remain in the dimension of task-related agency.

Autonomy of Process

Autonomy of performance indicates a crucial component of functional autonomy but does not account for the entirety of the phenomenon. There is at least another aspect that can be singled out. In the task-related dimension we are studying, two elements are of interest at first glance: autonomy in function execution as a whole and autonomy in selecting different courses of action to accomplish the given goal depending on various conditions. Let us focus on this second aspect by introducing the following case study:

Case Study 2. A thermostat is an automatic temperature regulation device; once set, it operates relatively autonomously. The recent car-sized Mars rover *Curiosity* is also an automatic device, and implements many complex functions. They both enjoy autonomy of performance: once activated, these devices are able to carry out given functions and accomplish given goals. Intuitively, however, it makes little sense to consider a thermostat and *Curiosity* as equally autonomous agents. Why is that? What further aspect of functional autonomy emerges here?

To get a closer look at this aspect of functional autonomy, let us start by considering a thermostat that regulates the temperature of a certain environment. The thermostat clearly enjoys autonomy of performance; that is, even if there were no other agents, it would still be capable of regulating the temperature. In fact, the thermostat was designed precisely to regulate temperature in the absence of human agents and, thus, it is expected to do so as specified.

Also, an extraordinarily complex machine like the rover *Curiosity* has been designed to perform some tasks autonomously, i.e., in the absence of human agents. Still, there seem to be relevant differences between *Curiosity* and a thermostat in terms of autonomy. These differences, moreover, appear to be reducible to something more than just a matter of degrees. Rather, they seem to point at the deeper level of logical discontinuity between the two artifacts.

In order to study this case and describe the second model of functional autonomy we need to consider the class of *procedures*, i.e., the ways an agent can count on to perform the task at hand. As stated in section 3, an agent *a* performs a task *t* according to procedure *p*. Procedures are sequences of elementary operations necessary for the completion of a task.

The relationship between an agent and the procedures it can adopt to carry out a given task describes a relevant dimension of functional autonomy. In the case of the thermostat, the completion of the task is uniquely determined by the procedure embedded in its design. In other words, the thermostat has *no degrees of freedom* with regard to the *way* of regulating the temperature. There is only one procedure *p* by which the thermostat regulates the temperature:

(4) thermostat, *p*: temperature control

Such rigidity is precisely what makes it odd to ascribe the same autonomy to the thermostat and *Curiosity*. The rover, as a more complex artificial agent, enjoys a much higher degree of freedom as to the way in which it pursues a certain task. Consider, for example, autonomous nav-

igation from point A to point B. If the rover has to reach a certain place on the Martian surface (let us call this task *mars*), it can compute which the best route is given the current environmental conditions on Mars, its own internal states, and other relevant factors. Let us express this as follows:

(5) ◊rover, p_1:*mars*
◊rover, p_2:*mars*
◊rover, p_3:*mars*
…

To sum up, the idea behind autonomy of process is that the autonomous agent can count on at least two possible procedures to obtain a specified task and has the ability to determine which one to apply given the specific conditions in which it works. Let us also refer to autonomy of process as autonomy$_2$. We will therefore say that:

(6) An agent *a* is autonomous$_2$ with respect to task *t* if and only if there are at least two procedures p_1 and p_2 the agent can select to perform task *t*:

$A_2(a, t)$ iff $\exists p_1 \exists p_2$ $(\Diamond a,p_1{:}t \,\&\, \Diamond a,p_2{:}t)$

Of course, this does not mean that the procedures are *indifferent*; one may clearly be better than the other. In this case, expectations and evaluations will revolve around the precision by which the technology selects the most effective procedure within those available given the circumstances.

Since autonomy of *process* is based on the availability of a range of possible procedures to accomplish a given task, it is plausible to think that it has a characterization: some agents are more autonomous$_2$ than others because of a higher degree of freedom in relation to the procedures they can select to carry out a specific task. In other words, this second aspect of functional autonomy captures the flexibility an autonomous agent exhibits in the execution of specified tasks. The more autonomy of process – or, similarly, the more flexibility – there is, the more efficiency and effectiveness there will be.

At the same time, more flexibility also implies more uncertainty in the relationship between users and artificial agents. Consider *Curiosity*'s navigation mission one last time. If we trust the technology, we can anticipate with some degree of confidence that *Curiosity* will be at the expected location on time; but it becomes rather hard to precisely anticipate each and every step the rover will undertake to get there. This seems to

be similar to what Johnson and Verdicchio have in mind when they characterize artificial agents' autonomy as the ability to "achieve a goal without having their course of action fully specified by a human programmer" (Johnson & Verdicchio, 2017, p. 577; see also Johnson & Verdicchio, 2019). Likewise, this feature is the main reason why, according to Matthias (2004), the use of artificial agents might lead to responsibility gaps.

Even in this second case, it remains important to distinguish between functional and full autonomy in task-related agency. The ways in which different procedures are available to artificial and human agents, and the ways in which procedure selection might occur, belong to widely different dimensions in the two cases, to the point that it would be risky to apply the same framework to both occurrences. In this regard as well it is necessary to thoroughly account not only for the similarities, but also for the differences that separate the two forms of autonomous agency.

Conclusion

The notion of autonomy has always played an important role in the philosophical understanding of human agency and, more generally, of the human condition. As automation progresses, new technologies exhibit forms of behavior that appear to be autonomous as well, thus adding new layers of meaning to the notion and, at the same time, introducing the possibility to compare human and artificial forms of autonomy. The autonomy artificial agents enjoy, however, presents some new and unique features. Clarifying these features is of utmost importance if we are to elaborate a rigorous philosophical account of artificial agency. In this paper, we have focused our attention on functional autonomy, and we have proposed a formal framework to analyze its most evident aspects. First, we have addressed functional autonomy as autonomy of performance, according to which an artificial agent can be said to be autonomous if and only if it is able to carry out a task without requiring other agents' intervention. Secondly, we have addressed functional autonomy as autonomy of process, according to which an artificial agent is autonomous if and only if it is able to select the procedure to accomplish a goal among a set of alternatives. We believe that these two models might help clarify the meaning of functional autonomy and, more generally, of artificial agency. Even though the results of the present analysis are limited, they show how important it is for the philosophy of robotics to shed light on the features of functional autonomy.

References

Amigoni, F., & Schiaffonati, V. (2005). Machine ethics and human ethics: A critical view. In *Proceedings of the AAAI 2005 Fall Symposium on Machine Ethics* (pp. 103-104).

Belnap, N., Perloff, M., & Xu, M. (2001). *Facing the Future*. Oxford University Press.

Bryson, J. J. (2010). Robots should be slaves. In Y. Wilks (Ed.), *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues* (pp. 63-74). John Benjamins Publishing Company.

Bryson, J. J., & Kime P. (2011). Just an artifact: Why machines are perceived as moral agents. In T. Walsh (Ed.), *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence* (pp. 1641-1646). AAAI Press.

Chiffi, D., & Pietarinen, A. V. (2020). Abductive inference within a pragmatic framework. *Synthese, 197*(6), 2507-2523.

Floridi, L. (2011). On the morality of artificial agents. In M. Anderson & S. L. Anderson (Eds.), *Machine Ethics* (pp. 184-212). Cambridge University Press.

Fossa, F. (2020). Etica funzionale. Considerazioni filosofiche sulla teoria dell'agire morale artificiale. *Filosofia, 65*, 91-106.

Franklin, S., & Graesser, A. (1996). Is it an Agent, or just a Program?: A Taxonomy for autonomous agents. In J. P. Müller, K. J. Wooldridge, & N. R. Jennings (Eds.), *Intelligent Agents III Agent Theories, Architectures, and Languages. ATAL 1996. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence): vol 1193* (pp. 21-35). Springer.

Gunkel, D. (2012). *The Machine Question. Critical Perspectives on AI, Robots, and Ethics*. MIT Press.

Johnson, D. G. (2011). Computer systems: moral entities but not moral agents. In M. Anderson, & S. L. Anderson (Eds.), *Machine Ethics* (pp. 168-1839). Cambridge University Press.

Johnson, D. G., & Verdicchio, M. (2017). Reframing AI discourse. *Minds and Machines, 27*(4), 575-590.

Johnson, D. G., & Verdicchio, M. (2019). AI, agency and responsibility: The VW fraud case and beyond. *Ai & Society, 34*(3), 639-647.

Lewis, D. (1973). *Counterfactuals*. Harvard University Press.

Maes, P. (1991). *Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back*. MIT Press.

Magnani, L. (2011). *Abduction, Reason and Science: Processes of Discovery and Explanation*. Springer.

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and information technology, 6*(3), 175-183.

Müller, V.C. (2021). Ethics of artificial intelligence and robotics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. (Summer 2021 ed.). Stanford University.

Russell, S.J., & Norvig, P. (2010). *Artificial Intelligence. A Modern Approach*. Prentice Hall.

Segerberg, K., Meyer, J. J., & Marcus, K. (2020). Ethics of artificial intelligence and robotics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. (Summer 2020 ed.). Stanford University.

Stalnaker, R. C., & Thomason, R. H. (1970). A semantic analysis of conditional logic 1. *Theoria, 36*(1), 23-42.

Steels, L. (1995). When are robots intelligent autonomous agents?. *Robotics and Autonomous systems, 15*(1-2), 3-9.

Sullins, J. P. (2006). When is a robot a moral agent. *Machine ethics, 6*, 23-30.

Tamburrini, G. (2020). *Etica delle macchine*. Carocci.

Verdicchio, M. (2017). An analysis of machine ethics from the perspective of autonomy. In T. Powers (Ed.), *Philosophy and Computing. Philosophical Studies Series, vol. 128* (pp. 179-191). Springer.

Wheeler, M. (2020). Autonomy. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *Oxford Handbook of Ethics of AI* (pp. 343-358). Oxford University Press.

Williamson, T. (2016). Abductive philosophy. *The Philosophical Forum, 47*(3-4), 263-280.

Winfield, A. (2012). *Robotics. A Very Short Introduction*. Oxford University Press.

Woods, J. (2013). *Errors of Reasoning. Naturalizing the Logic of Inference*. College Publications.

4. Human Experience and Robotic Experience

A Reciprocal Exchange of Perspectives

C. Mazzola, S. Incao, F. Rea, A. Sciutti, M. Marassi

ABSTRACT

The basis of a humane approach to others is the authentic comprehension of another subject. As humans, we can achieve this understanding well, and the reason lies in how we experience the world around us and other subjects in it. The development of robots capable of socially interacting and helping humans is progressing, even though they are still far from reaching an autonomous comprehension of others' intentions, emotions, and feelings. In this sense, the humane approach may be addressed in robotics through the concept of experience. In a reciprocal exchange of perspectives, the core elements and the structure of human experience are investigated in this chapter together with how the idea of experience has been implemented in robots. The embodied Self and the relationship with other subjects form the pivot of any human experience and are suggested to be the basis for the emergence of a novel cognitive-experiential structure in robots. Conversely, the possibility to develop a robot with a primitive sense of Self raises questions about the nature of human experience and the impact such technologies have on it.

The Experience as a Core Element of a Novel Interdisciplinary Exchange

Cognitive robotics inspired by human beings

In robotics, a specific branch aims to endow robots with models inspired by the most advanced cognitive agent known: the human being. The objective of Cognitive Robotics is to develop and unify systems of perception, action, memory and learning that might provide robots with higher autonomy and flexibility when interacting with the environment (Vernon, 2014; Sandini et al., 2019). One of the merits of this branch is to unite the practical implementation of cognitive and neuroscience research outcomes on a robotics platform equipped with a physical artificial body. This approach may become even more intriguing, considering humans more as subjects of experience than as cognitive systems. Indeed, to have an experience or to experience something is part of

---

C.M. and S.I. contributed equally to this research.

our daily life as human beings. We have experiences about the world when we taste chocolate, touch the frozen snow, or look at the colorful feathers of a tropical parrot. However, we can also experience our painful muscles after a long swim or butterflies in the stomach before presenting at an important conference. These experiences, exteroceptive, proprioceptive, and interoceptive respectively, are inherently tied to the body and allow us to interact in the environment and with others.

Robotic experience

When dealing with artificial robotic systems, the term 'experience' acquires a specific connotation. Robotic platforms are endowed with multiple kinds of sensors through which the robot retrieves information from the environment. Moreover, robots can also continuously collect information about their inner state: for example, the position of their joints and the velocity of their movements. All these data can be stored in memory and processed by algorithms targeted to make the robot situated in the environment, to optimally plan the robot's actions or train the robot on a specific task.

For specific applications, where the context and task are well defined and fixed – such as, for example, traditional automotive production – robots pre-programmed to reproduce the same operations accurately may represent an effective solution. Conversely, a robot meant to navigate everyday life in society needs to be endowed with the ability to learn and adapt to the ever-changing environment and the unpredictability of human beings (Vernon et al., 2016). For this reason, the application of machine learning algorithms to robots represents a massive field of research with significant results, although highly task-dependent. For example, a particular technique used to train robots is Reinforcement Learning. Here, positive and negative rewards received in continuous interactions with the world reinforce the learning process. In this way, the robot can "learn from its experiences", i.e., find and learn patterns in the environment and any related information regarding the world around it.

Experience from a first-person perspective

The sense of the word 'experience', as targeted by robotics, is in line with the operational aspect of human experience. Indeed, the methods we employ to face our daily interactions with the world involve our body in a combination of perception and action (Varela et al., 2016). Since childhood, our bodily experience has been the background of every learning process (Martin & Schwartz, 2005; Falck-Ytter et al., 2006). The

experience of the world is structured by methods and patterns that we memorize and repeatedly reuse, adapting them to any occurrence. However, even though this operational aspect of experience is implicit in the everyday life of human beings, the meaning of the term 'experience' in humans invokes something even more articulated as a concept. Indeed, any experience refers to a specific subject possessing a personal or subjective character that concerns the first-person perspective of the individual. Every experience is therefore typical for the person to whom it belongs. It originates as related to the intrinsic motivations of that subject, colored by their affective states, intertwined with other experiences, structured in their moral beliefs.

Experience and meaning

In this sense, human experience is what allows us to interpret and give meaning to the world from a specific and unique perspective, so that every object (which might be something in the environment or another person) is perceived with reference to ourselves: our motivations, expectations, emotions, personality, memory, body. Therefore, whereas robots can already detect, store and learn information from the environment to solve a task, this ability is still far from the complexity of the human experience of the world. Furthermore, cognitive skills are necessary to devise meaning from the information processing of autonomous agents. The information coming from the environment through the senses gains meaning from motivations connected to the organism's maintenance of balance. Damasio explains this process through the concept of homeostasis (Man & Damasio, 2019). The motivation behind human actions and behaviors cannot always be traced back solely to the need for survival. In fact, the scale of values we use to act and behave implies affective, instinctual, and moral incentives. Reference to the Self and to the maintenance of a sort of equilibrium is therefore crucial for human beings. Through this constant contact with their individual point of view, human beings can overcome the limitations of pure data received from the outside, by interpreting them in light of motivations, values, affections, memories and moral beliefs. Moreover, it is only through this constant allusion to an individual coherence that humans can establish a hierarchy in the flow of experiences, which otherwise would be a set of equivalent information.

Promotion of the dialogue

Albeit linked to the specificity of human experience (lacking in robots), dialogue around what makes an experience possible might be highly

relevant in both the robotic and philosophical fields. Investigating the structure of human experience might inspire novel approaches toward the improvement of robotic autonomy and social abilities (Gaggioli et al., 2021). On the other hand, discussion about the advances and potential developments of artificial embodied systems might promote a deeper investigation of the nature of human experience and the impact such technologies have on it.

A Philosophical Investigation of Experience: The Priority of Interpretation

Definition of experience

As a preliminary and broad definition, *experience* outlines the way humans internalize the reality they live in. The word stems from the Latin *ex-periri*, which means to go through a trial, to attempt something. Therefore, it describes a form of knowledge concerning a consistent and repetitive situation that produces the conception and the understanding of a State of Things, the direct relation of the subjects to the otherness they are affected by, the ability to remember and organize impressions, the possibility to broaden knowledge through a verification process.

Experience as repetition

In all its forms, experience implies a series of events that appear to the subject who experiences them. The temporal sequence in which they appear includes the presence of elements that are recognized as analogous. Hence, the primary feature that describes experience is repetition. Then, backward reflection about the flow of different experiences allows the uniqueness of each event to be identified.

In fact, it might appear paradoxical that such uniqueness becomes evident only through the constant repetition of a phenomenon. However, it is only in this way that its nature may become known and its conditions of possibility identifiable. Moreover, only through this process it is possible to anticipate a novel event and verify its evidence. That being the case, the repetition of events might produce a twofold effect. The first is the prediction of a phenomenon not yet experienced but analogous to previous ones. The second is the identification of a uniformity of phenomena that can be formulated as a general law. More precisely, on the one hand prediction relies on repetition, while on the other hand, uniformity – once defined as a general law – exceeds iteration and is irrespective of temporality.

Theory and observed data in science

Reflection on the connotation of experience in science might deepen the relationship between a single experienced event and the constant general law that expresses its meaning. The parallel in the scientific method would be the structure of the experiment on the binomial of observed data and scientific theory. As Kant pointed out, explicitly referring to the Newtonian physics and Galilean astronomy of his time, no hypothesis could be deduced from observation (Kant, 1787). Therefore, the experiment is not a step towards creating a theory; it does not genetically precede the idea, but is rather the expression of fundamental questions that humans pose to nature through theoretical hypotheses. Kant saw that the history of science confuted the emergence of a theory as a derivation from repeated observations. As he advocated, knowledge does not emerge from facts and phenomena but from *a priori* forms that build and order them. Moreover, as Popper expounded, the necessity and universality of theories do not arise from experiments. Experiments are instead suggested and interpreted by theories (Popper, 1962).

The priority of lifeworld

In light of this comparison with science, it is essential to explore a distinctive aspect of human experience by moving from an opposition. On the one hand, scientific experience is characterized by the pursuit of objectivity achieved through a method. On the other hand, human experience is marked by historicity, by the novelty manifested in the timing of events, and by the unifying nature of consciousness that collects them as a teleologically organized continuity of memory, intentionality and forethought.

Experience, as intended in the scientific method, needs to be confirmed and validated by repetition: *ubi non reperitur instantia contradictoria* (Bacon, 1623; Stuart Mill, 1882). However, this connotation of experience does not imply the idea of historicity: it is far from being an existential experience. Before any process of abstraction and generalization, a subject is situated in the lifeworld. The existential experience does not proceed by progressive abstractions, since any phenomenon appears against a background of subjective interpretation. Therefore, the process of experience does not move from the observation of a fact toward its generalization. Instead, the perception of a single event emerges from motivations, personal experiences and bodily dispositions that interpret its meaning and move toward the consequent teleologically oriented action execution.

The meaning of denial

There is another point that demonstrates the difference between experience as a process of generalization/abstraction and experience as an interpretative process of events: this is the different meaning of denial in science and life. In scientific experiments, if the hypothesis gets disconfirmed, the error can be amended by formulating a new theory. In human experience, denial is called blame, error or evil i.e., something which is not entirely identifiable with the universality of a concept since it refers deeply to the inner state of the subject. Thus, human experience takes place in single events interpreted in light of the subject's history: previous delusions and faults, current motivations, and prospects for the future. It is far from being a repetitive process in which every case is reduced to the universality of its generalization.

Bodily experience and interpretation

The cornerstone of every human experience is the body. The body is indeed the only means through which a subject is open to a relationship with the environment (Merleau-Ponty, 1945). However, it is essential not to consider it a mere sensory apparatus that allows data coming from the senses to enter the conscience. The perceived object is never neutral for the subject, it is never experienced as mere information. Instead, it is always perceived relative to subjectivity and directly interpreted by the body. In this sense, the body is not inanimate and lifeless. It is rather living and lived. Lived because it collects, like tree bark, the imprints that the flow of time leaves on it. Living because it continuously receives sensory stimuli from the external environment that strikes it in a specific – its own specific – way. Perception is thus the bodily relationship between the subject and the object, the primary form of human experience where the external world appears to the subject as already interpreted, meaningful, relative to the Self.

Self-Awareness and Bodily Experience

Experience of the body

Having a body and *being a body*. These two expressions disclose the twofold nature of the body: that of being simultaneously experienced as an object and as a subject. As a matter of fact, the body differs from all other objects present in the world. It is always part of us, and we cannot

part with it or live without it. The expression *having a body* highlights the body's role as the origin of our contact with the world. The body we possess can sense the world and is the means by which we interact with the sensory world.

For this reason, the idea of *having* a body alludes to the experience of an entity we can feel, sense, touch, think about, conceptualize, and take as the object of our thoughts. The uniqueness of the object body is that it is experienced in first person and therefore becomes part of a subjective perspective. Coupled with the feeling of *having a body* is the experience of *being that body*, i.e., the awareness of it as the condition of possibility for any imaginable experience. It is myself as the body I am that lives and interacts, making any experience possible.

Body schema

Artificial bodies are supposed to integrate with the algorithmic *brain* of robots in performing tasks that humans usually carry out. A specific field of robotics research is focused on addressing this integration (Vernon et al., 2016), supporting the view that cognition is "deeply dependent upon the characteristics of the physical body of an agent" (Wilson & Foglia, 2011). Doing things like walking across a room avoiding obstacles is a simple action for an adult human being, but it may be slightly more difficult for a child and even more complex for a robot. A lot depends upon the extent to which our body schema has grown accustomed to the world outside. Many authors have used the concept of body schema since the beginning of the last century, but it has sometimes suffered from the lack of any conclusive definition. However, as Gallagher (1986) points out, the origin of ambiguities lies in the doubts regarding the conscious or unconscious nature of the body schema besides the body image. For this contribution, it is helpful to follow Gallagher's conclusion. He explains that the body schema, in contrast to the body image, is not a conscious percept through which we think about our body as the object of our thoughts. The body schema is, rather, the non-conscious constant responsible for our body's operative performance in the environment. When we are about to cross a road or receive a ball during a volleyball match, the postures we assume and the movements we perform are caused by the organization of the body schema. It contains information such as the space occupied by the body or the arm's length, which determines the estimations necessary to grasp or avoid objects and many other details about the functional and operative performance of the body within the world. The concept of body schema has introduced a modality through which human beings build a sort of model or

map of the world outside that is strictly linked to the experience of their own body. Indeed, this connection results in a direct and mainly unconscious contact with the environment.

The body and the Self interpret the world

The interactions with the world that form specific human experiences are characteristic to each individual not only because the shape of human bodies differs, but because the experience of the world differs considerably between distinct human beings. The human body, and consequently the body schema, are inherently tied to the subjective perspective of everyone's specific experience. Human beings interpret the world based on a personal horizon of experience, which is the sum of motivations, beliefs, intentions, and actions towards reality. Every new experience from the outside takes shape internally, always in reference to the flow of our previous experiences. This background of subjective participation in the world refers to the concept of Self.

Artificial Self

The progressive resemblance with humans that artificial embodied systems are pioneering is not only at the level of physical appearance but also in terms of behavior during interaction with the environment. From this perspective, the individual unity that we experience as human beings, the Self, has been recognized as a core element, if robots are to gain autonomy and adaptivity during interactions. As a result, several studies on cognitive robotics are looking at the possibility of building robots with a sense of Self. This element is intended to be included in a cognitive architecture, a computational model of the structure of a human mind connecting different modules such as memory, perception, decision-making, action and learning, so as to confer functional autonomy to the robot. In particular, the question is whether the Self should be a fixed module, a sort of metacognitive unity in which information is conveyed and analyzed, or if the Self should rather be seen as an emergent feature that progressively develops and rises through the learning of new data.

The Self as a module or an emergent element

Since the early 2000s, robotics research has attempted to improve the cognitive and relational capabilities of robots. Self-awareness has been

identified as the key element to develop such capabilities (Lee, 2020). Many studies have attempted to define and build a model of the Self to be added to the cognitive architecture.

In 2005, Kawamura et al. (2005) attempted to develop a robot with a sense of Self by using a cognitive architecture with three memory systems that are repeatedly accessed to monitor the robot's representations of its internal states and the progress of the task required. In their experiment, the robot was able to suspend the task it was carrying out to pay attention to a new stimulus identified as more urgent.

Also, Novianto and Williams (2009) have defined Self-awareness in robots to be characterized by the ability to redirect attention towards internal states. Here the focus is on the attentional process that is "the allocation of perceptual resources to analyze a subset of the surrounding world to the detriment of others" (Ferreira & Dias, 2014). Following Novianto & Williams (2009), the attentional process should be directed towards internal states to develop self-awareness in a robot.

In contrast, in Birlo and Tapus (2011) the Self is configured as an independent unit that monitors different memory modules, namely *buffers*, and decides which one requires attention. Additionally, in Chatila et al. (2018) the cognitive architecture is designed in modules. Each one of these modules is a decision-making system that implements one way of producing actions. In addition to this system, there is a meta-element called *meta-controller* responsible for analyzing every decision-making process and selecting the one in charge of controlling the robot at any given time.

All these studies show that self-awareness has been considered a crucial element for robots to interact naturally with humans.

Furthermore, other studies, inspired by developmental psychology, have attempted to let the Self emerge with unsupervised learning mechanisms. Starting from data acquired from the environment, the robot could generate emergent representations. The concept of *emergence* emphasizes the capability of a system to show properties or abilities that are not intrinsic to its components. Hoffmann et al. (2018), for example, proposed an approach to study how the humanoid robot iCub could form a representation of its body surface by receiving several stimuli on its artificial skin with capacitive sensors. Here, the topographic representation is not pre-scripted but emerges from the robot's interaction with the surrounding environment. Furthermore, Lanillos and Cheng (2018) developed a perceptual computational model for multisensory robots to derive their body configuration through sensory information.

These implementations of Self modules are more focused on developing a concept of self-awareness as direct knowledge of the internal states. The approaches inspired by developmental psychology are, in-

stead, closer to the embodied cognitive processes humans experience in the first stages of their lives, when a primary and unconscious sense of Self emerges in the womb. Both of these approaches, addressing the topic from different perspectives, may be instrumental in providing efficient and operative solutions to improve the adaptivity and autonomy of robots.

Perception and Intersubjectivity

Perceiving others through our body

A noticeable difference lies between perceiving the environment and perceiving other subjects in it. In both cases, our senses receive numerous impressions, but the body participates in perception differently. A deep connection exists between others' bodies and one's own so that the body is doubly involved. Others' movements, actions and even bodily reactions are received with sight, hearing and touch but are experienced by passing through one's own body schema, reflecting and interpreting the other's bodily experience. This is also suggested by the function of the mirror neuron system, a cerebral organization of neurons that activates both when one is performing an action or when one is merely seeing others performing the same action (Rizzolatti et al., 1996; Rizzolatti & Craighero, 2004).

Twofold maturation of Self and perception of others

After all, the strict ties with others' bodies are evident since the prenatal stage, when the fetus grows, is fed and takes oxygen within the maternal womb. This strong relationship with the mother remains similarly intense through the first months of life. From birth, the newborn understands that other persons are 'like me'. Only from the ninth month they start considering others as other intentional agents, an understanding they will gradually develop over the years (Tomasello, 1999). Finally, in early childhood, this process reaches a critical step. The child starts distinguishing their mental states from those of others, avoiding the mere egocentric attribution (Moll & Meltzoff, 2011; Tomasello, 1999). During this process of distinction, the Self emerges. The Self is not indeed a static, already formed structure since birth. It instead comes to consciousness and develops in a continuous comparison with others, that are gradually recognized as other different selves having other diverse experiences. Such a twofold maturation – the emergence of the Self and

recognition of other selves – persists throughout life and shows that, whilst the Self develops in relation to others, others are perceived and their experiences interpreted based on one's own Self. Therefore, the perception of others emerges from an analogical process based on oneself, one's own experiences, and one's own body.

Primitive skills of understanding others in humans and robots

Since social interaction is one of the ten main challenges in robotics (Yang et al., 2018), cognitive and developmental robotics aims to provide robots with the social ability to understand other agents' states (Sciutti & Sandini, 2017). In this direction, robots have been endowed with the primitive social skills infants acquire during the first months or even weeks of life, such as the predisposition to focus on biologically moving objects rather than non-biological ones (Simion et al., 2008; Vignolo et al., 2017). Also, the ability to discriminate an averted from a direct gaze, and the preference for the latter (Farroni et al., 2002), as well as to detect the direction of the gaze, have been implemented in artificial embodied systems (for example, see Schillingmann & Nagai, 2015; Palinko et al., 2016).

Development of Social Skills in Humans and Robots

All these abilities are at the basis of higher cognitive and social skills along the developmental process. Gaze direction detection enables joint attention mechanisms, which underlie perspective-taking and higher forms of understanding others' intentions (Tomasello et al., 2005; Moll & Meltzoff, 2011). Perspective-taking is indeed the ability to grasp the visuospatial perspective of another person that is different from the egocentric one, and gather what is seen by them, and how (Salatas & Flavell, 1976; Flavell et al., 1981). Such an ability seems to occur both as a mentalizing process and as an implicit and automatic mechanism (Surtees & Apperly, 2012). In robotics, such skill has been developed on artificial platforms to improve collaboration with other agents (Trafton et al., 2005; Zhao et al., 2016) and to enable robots to better map the environment in relation to their human partners (Fischer & Demiris, 2016). Furthermore, in infancy, the process matures up to the development of a Theory of Mind (Baron-Cohen, 1995, 2000) that allows children to understand the beliefs, intentions, and desires of other individuals. This same cognitive skill was found to be fundamental for robots involved in social interactions at the dawn of modern robotics (Scassellati, 2002)

and was later implemented in social robots as an adaptive skill (Bianco & Ognibene, 2019).

The body is a means to interpret others

The child's developmental process shows the fundamental role of sociality on the emergence of the child's Self. The Self grows in a continuous comparison with others so that the social context, which will see the child acquiring language skills and symbolic thought, starts earlier, shaping and building the infant's Self. Indeed, the process of distinction from the mother and the progressive formation of others' mind consciousness manifest the body's role in developing self-other distinction. As the neural evidence of mirror neurons also suggests (Rizzolatti & Craighero, 2004), the function of the body, and more precisely the body schema, also persists into adulthood. When perceiving other persons, they are interpreted through one's own body as a lens to experience the environment and other agents. Hence, one's own body is the means to interpret what is perceived of others. Personal body schema, previous experiences, affective states, instincts, motivations lie in the body and are essential to comprehending others, which are experienced and interpreted through one's own embodied Self. Therefore, since the same ability to understand others would be crucial also for robots collaborating with humans (Sciutti & Sandini, 2017), robots should be provided with a sense of embodied Self (Prescott & Camilleri, 2019), serving as a pivot to understand others.

The emergence of self-other distinction skills in robotics

The tight connection between one's own body and perception of others has been explored in cognitive robotics with interesting results. A fundamental ability allowing the understanding of others is the self-other distinction. Zhang and Nagai (2018) proposed a computational model for learning the internal bodily states of the robot that brings together proprioceptive feedback from the robot's joints and visual feedback from its cameras. In this way, the robot can predict its own proprioceptive feedback starting from observing itself, an ability that would simulate the imaginary body schema. Moreover, the skill of self-other distinction can emerge from the robot learning it can visually distinguish its own actions from those of others. Following the same inspiration, Lanillos et al. (2020) developed a learning algorithm that allows the robot to recognize itself in a mirror by recognizing some simple actions generated by itself, with the consequent ability to distinguish its own movements from the ones generated by others.

Learning to understand others' intentions in robotics

The ability to understand the intentions of a partner and anticipate their movements is essential for collaboration. The cognitive architecture Hammer (Demiris & Khadhouri, 2006) was inspired by the mirror neuron system to provide robots with the ability to understand others' actions. In that architecture, the robot uses its motor control system, both for executing actions and for perceiving the actions performed by other agents.

Moreover, Vinanzi et al. (2019) proposed an artificial cognitive architecture for intention prediction. The idea, achieved with an unsupervised learning model, is based on detecting others' skeletons during their action and inferring the subsequent postures up to the end of the action. Copete et al. (2016) introduced a computational model for action prediction based on the co-development of action prediction and action production. The sensorimotor information gathered during action production, which is missing in others' action observation (tactile and motor feedback), were reconstructed based on the information learnt in action production to facilitate the other's action prediction.

Cognitive robotics research is therefore implementing social abilities in robots, which are inspired by the developmental process of children. As evidenced in the paragraphs above, robots have already been endowed with some of these abilities and can learn a number of social interaction skills. Perhaps, what may be missing is a stronger focus on the Self as a mean to understanding others, an element which is at the basis of all human experience.

Conclusive Remarks

The basis of a humane approach to others is the authentic comprehension of another subject. Every behavior, inspired by this motivation, proceeds from a prior understanding of what others are feeling, their emotions and intentions. A fundamental requirement for a humane robot is, therefore, to understand others (Sandini & Sciutti, 2018). As humans, we can achieve that understanding only through our own embodied Self: we comprehend others starting from ourselves – an ability which infants develop over the years. Technological advances have already provided robots with many social skills. Attempts to provide robots with a primitive sense of Self and distinction from others have been conducted by leveraging machine learning techniques and starting from bodily interactions with the world around them.

To continue in this direction, we believe that the idea of experience outlined in this chapter needs also to be considered for robots, some-

thing that can be obtained only through a multidisciplinary dialogue. Cognitive robotics might indeed receive inspiration from the human way of experiencing the world and others. The concept of an embodied Self as a pivot of experience is cardinal and evidences the constitutional relationship between what is experienced and the subject of experience. As artificial cognitive systems, robots should therefore be developed in light of the concepts of Self and experience. The Self is not reduced to the ability of being aware of one's own internal states. It can reach a conscious level with reflection, but it is also the subject's flow of experiences. In this sense, it can be viewed as the reference and interpreting medium for any possible experience appearing in the flow. To develop a subjective experience, then, robotics needs to undergo a mindset shift: from the integration of information to the interpretation of such information on the basis of the embodied Self. Interpretation means that every piece of information is read as related to the embodied Self: its body schema, motivations, affective states, previous perceptions, and personality. Such an Artificial Self would provide robots with a cognitive-experiential structure inspired by humans and would make them more autonomous in learning, and more able to understand other agents and establish a commonality with them.

On the other hand, there are also prolific perspectives on this topic linked to the philosophical thought behind research into the specific nature of human cognition and experience. The building of robots that can socially interact with humans, learn from the environment, and develop a primitive sense of Self raises questions about human nature and its difference from artificial systems. It is also crucial to explore the impact that novel artificial systems may have on humans. Indeed, social robots may mediate and influence the way humans interact with the world around them, as every technological device does (Ihde, 1990).

From this perspective, any research into human experience and the development of robots with a sense of Self and capable of experiencing others, should cooperate and proceed in parallel with the aim of a human-centric technology.

References

Bacon, F. (1623). *De Augmentis scientiarum.*

Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind.* MIT Press/Bradford Books.

Baron-Cohen, S. (2000). Theory of mind and autism: A review. *International Review of Research in Mental Retardation, 23*, 169-184. https://doi.org/10.1016/s0074-7750(00)80010-5.

Bianco, F., & Ognibene, D. (2019). Functional advantages of an adaptive Theory of Mind for robotics: A review of current architectures. *2019 11th Computer Science and Electronic Engineering (CEEC)*, 139-143. https://doi.org/10.1109/CEEC47804.2019.8974334.

Birlo, M., & Tapus, A. (2011). The crucial role of robot self-awareness in HRI. *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 115-116. https://doi.org/10.1145/1957656.1957688.

Chatila, R., Renaudo, E., Andries, M., Chavez-Garcia, R. O., Luce-Vayrac, P., Gottstein, R., Alami, R., Clodic, A., Devin, S., Girard, B., & Khamassi, M. (2018). Toward self-aware robots. *Frontiers Robotics AI*, *5*(AUG). https://doi.org/10.3389/frobt.2018.00088.

Copete, J. L., Nagai, Y., & Asada, M. (2016). Motor development facilitates the prediction of others' actions through sensorimotor predictive learning. *2016 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, 223-229. https://doi.org/10.1109/DEVLRN.2016.7846823.

Demiris, Y., & Khadhouri, B. (2006). Hierarchical attentive multiple models for execution and recognition of actions. *Robotics and Autonomous Systems*, *54*(5), 361-369. https://doi.org/10.1016/j.robot.2006.02.003.

Falck-Ytter, T., Gredebäck, G., & von Hofsten, C. (2006). Infants predict other people's action goals. *Nature Neuroscience*, *9*(7), 878-879. https://doi.org/10.1038/nn1729.

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *PNAS*, *99*(14), 9602-9605. https://doi.org/10.1073/pnas.152159999.

Ferreira, J. F., & Dias, J. (2014). Attentional mechanisms for socially interactive robots - A survey. *IEEE Transactions on Autonomous Mental Development*, *6*(2), 110-125. https://doi.org/10.1109/TAMD.2014.2303072.

Fischer, T., & Demiris, Y. (2016). Markerless perspective taking for humanoid robots in unconstrained environments. *Proceedings - IEEE International Conference on Robotics and Automation*, *2016-June*, 3309-3316. https://doi.org/10.1109/ICRA.2016.7487504.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1-Level 2 distinction. *Developmental Psychology*, *17*(1), 99-103. https://doi.org/10.1037/0012-1649.17.1.99.

Gaggioli, A., Chirico, A., Di Lernia, D., Maggioni, M. A., Malighetti, C., Manzi, F., Marchetti, A., Massaro, D., Rea, F., Rossignoli, D., Sandini, G., Villani, D., Wiederhold, B. K., Riva, G., & Sciutti, A. (2021). Machines like us and people like you: Toward human-robot shared experience. *Cyberpsychology, Behavior, and Social Networking*, *24*(5), 357-361. https://doi.org/10.1089/cyber.2021.29216.aga

Gallagher, S. (1986). Body image and body schema: A conceptual clarification. *The Journal of Mind and Behavior*, *7*(4), 541-554.

Hoffmann, M., Straka, Z., Farkas, I., Vavrecka, M., & Metta, G. (2018). Robotic homunculus: Learning of artificial skin representation in a humanoid robot motivated

by primary somatosensory cortex. *IEEE Transactions on Cognitive and Developmental Systems*, *10*(2), 163-176. https://doi.org/10.1109/TCDS.2017.2649225.

Ihde, D. (1990). *Technology and the Lifeworld. From Garden to Earth*. Indiana University Press.

Kant, I. (1787). *Kritik der reinen Vernunft* (II).

Kawamura, K., Dodd, W., Ratanaswasd, P., & Gutierrez, R. A. (2005). Development of a robot with a sense of self. *Proceedings of IEEE International Symposium on Computational Intelligence in Robotics and Automation, CIRA*, 211-217. https://doi.org/10.1109/cira.2005.1554279.

Lanillos, P., & Cheng, G. (2018). Adaptive robot body learning and estimation through predictive coding. *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 4083-4090. https://doi.org/10.1109/IROS.2018.8593684.

Lanillos, P., Pages, J., & Cheng, G. (2020). Robot self/other distinction: active inference meets neural networks learning in a mirror. *ECAI 2020: Proceedings of the 24th European Conference on Artificial Intelligence*, 2410-2416. https://doi.org/10.3233/FAIA200372.

Lee, M. H. (2020). *How to Grow a Robot: Developing Human-friendly, Social AI*. MIT Press.

Man, K., & Damasio, A. (2019). Homeostasis and soft robotics in the design of feeling machines. *Nature Machine Intelligence*, *1*(10), 446-452. https://doi.org/10.1038/s42256-019-0103-7.

Martin, T., & Schwartz, D. L. (2005). Physically distributed learning: Adapting and reinterpreting physical environments in the development of fraction concepts. *Cognitive Science*, *29*(4), 587-625. https://doi.org/10.1207/s15516709cog0000_15.

Merleau-Ponty, M. (1945). *Phénoménologie de la perception*. Librairie Gallimard.

Moll, H., & Meltzoff, A. N. (2011). Joint attention as the fundamental perspective in understanding perspectives. In *Joint attention: New developments in psychology, philosophy of mind*.

Novianto, R., & Williams, M. A. (2009). The role of attention in robot self-awareness. *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*, 1047-1053. https://doi.org/10.1109/ROMAN.2009.5326155.

Palinko, O., Rea, F., Sandini, G., & Sciutti, A. (2016). A Robot reading human gaze: Why eye tracking is better than head tracking for human-robot collaboration. *IEEE International Conference on Intelligent Robots and Systems*, *2016-Novem*, 5048-5054. https://doi.org/10.1109/IROS.2016.7759741.

Popper, K. (1962). *Conjectures and Refutations: The Growth of Scientific Knowledge*. Routledge.

Prescott, T. J., & Camilleri, D. (2019). The synthetic psychology of the self. In *Cognitive Architectures* (pp. 85-104). Springer.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*. 27, 169-192. https://doi.org/10.1146/annurev.neuro.27.070203.144230.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*(2), 131-141. https://doi.org/10.1016/0926-6410(95)00038-0.

Salatas, H., & Flavell, J. H. (1976). Development of two components of knowledge. *Child Development*, *47*(1), 103–109. https://doi.org/10.2307/1128288.

Sandini, G., & Sciutti, A. (2018). Humane Robots - from robots with a humanoid body to robots with an anthropomorphic mind. *ACM Transactions on Human-Robot Interaction*, *7*(1), 5-8. https://doi.org/10.1145/3208954.

Sandini, G., Sciutti, A., & Vernon, D. (2019). Cognitive robotics (in press). In M. H. Ang, O. Khatib, & B. Siciliano (Eds.), *Encyclopedia of Robotics*. Springer.

Scassellati, B. (2002). Theory of Mind for a humanoid robot. *Autonomous Robots*, *12*, 13-24. https://doi.org/10.1023/A:1013298507114.

Schillingmann, L., & Nagai, Y. (2015). Yet another gaze detector: An embodied calibration free system for the iCub robot. *IEEE-RAS International Conference on Humanoid Robots*, *2015-December*, 8-13. https://doi.org/10.1109/HUMANOIDS.2015.7363515.

Sciutti, A., & Sandini, G. (2017). Interacting with robots to investigate the bases of social interaction. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *25*(12), 2295-2304. https://doi.org/10.1109/TNSRE.2017.2753879.

Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *PNAS*, *105*(2), 809-813. https://doi.org/10.1073/pnas.0707021105.

Stuart Mill, J. (1882). *A System of Logic, Ratiocinative and Inductive 1843* (VIII). Harper & Brothers.

Surtees, A. D. R., & Apperly, I. A. (2012). Egocentrism and automatic perspective taking in children and adults. *Child Development*, *83*(2), 452-460. https://doi.org/10.1111/j.1467-8624.2011.01730.x.

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Harvard University Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675-691. https://doi.org/10.1017/S0140525X05000129.

Trafton, J. G., Cassimatis, N. L., Bugajska, M. D., Brock, D. P., Mintz, F. E., & Schultz, A. C. (2005). Enabling effective human-robot interaction using perspective-taking in robots. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *35*(4), 460-470. https://doi.org/10.1109/TSMCA.2005.850592.

Varela, F. J., Thompson, E., & Rosch, E. (2016). *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.

Vernon, D. (2014). *Artificial Cognitive Systems: A Primer*. MIT Press.

Vernon, D., von Hofsten, C., & Fadiga, L. (2016). Desiderata for developmental cognitive architectures. *Biologically Inspired Cognitive Architectures*, *18*, 116-127. https://doi.org/10.1016/j.bica.2016.10.004.

Vignolo, A., Noceti, N., Rea, F., Sciutti, A., Odone, F., & Sandini, G. (2017). Detecting biological motion for human-robot interaction: A link between perception and action. *Frontiers Robotics AI*, *4*(JUN). https://doi.org/10.3389/frobt.2017.00014.

Vinanzi, S., Goerick, C., & Cangelosi, A. (2019). Mindreading for robots: Predicting intentions via dynamical clustering of human postures. *2019 Joint IEEE 9th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, 272-277. https://doi.org/10.1109/DEVLRN.2019.8850698.

Wilson, R., & Foglia, L. (2011). Embodied cognition. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Summer 2021 Edition)*. https://plato.stanford.edu/archives/sum2021/entries/embodied-cognition.

Yang, G.-Z., Bellingham, J., Dupont, P. E., Fischer, P., Floridi, L., Full, R., Jacobstein, N., Kumar, V., McNutt, M., Merrifield, R., Nelson, B. J., Scassellati, B., Taddeo, M., Taylor, R., Veloso, M., Wang, Z. L., & Wood, R. (2018). The grand challenges of Science Robotics. *Science Robotics, 3*, 31. https://doi.org/10.1126/scirobotics.aar7650.

Zhang, Y., & Nagai, Y. (2018). Proprioceptive feedback plays a key role in self-other differentiation. *2018 Joint IEEE 8th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, 133-138. https://doi.org/10.1109/DEVLRN.2018.8761042.

Zhao, X., Cusimano, C., & Malle, B. F. (2016). Do people spontaneously take a robot's visual perspective? *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 335-342. https://doi.org/10.1109/HRI.2016.7451770.

SECTION 2

Exploring Human-Robot Interaction

The Influence of Cognitive and Affective Processes

1. Users' Affective and Cognitive Responses to Humanoid Robots in Different Expertise Service Contexts

Y. Jung, E. Cho, S. Kim

ABSTRACT

The uncanny valley (UCV) model is an influential human-robot interaction theory that explains the relationship between the resemblance that robots have to humans and attitudes toward robots. Despite its extraordinary worth, this model remains untested in certain respects. One current limitation is that the model has only been examined in general or context-free situations. Given that humanoids function in the world beyond laboratories, investigating the UCV in specific and actual situations is critical. Additionally, few studies have examined the impact of affective responses presented in the UCV to other appraisals of humanoids. To address these issues, this study explored affective and cognitive responses to humanoids in specific service situations. In particular, we examined the effect of affective responses on trust, which is regarded as a critical cognitive factor influencing technology adoption, in two service contexts: hotel reception (low expertise) and tutoring (high expertise). By providing a richer understanding of human both affective and cognitive reactions to humanoids, our findings expand the UCV theory and ultimately contribute to research regarding user adoption of service robots.

Introduction

Despite the diffusion of service robots, some practical warning signs have emerged. Many stores and companies that used Softbank's Peppers robots to interact with consumers have decided to discontinue their use because of customer resistance (Nichols, 2018). While robots' data management and multilingual capabilities are attractive to consum-

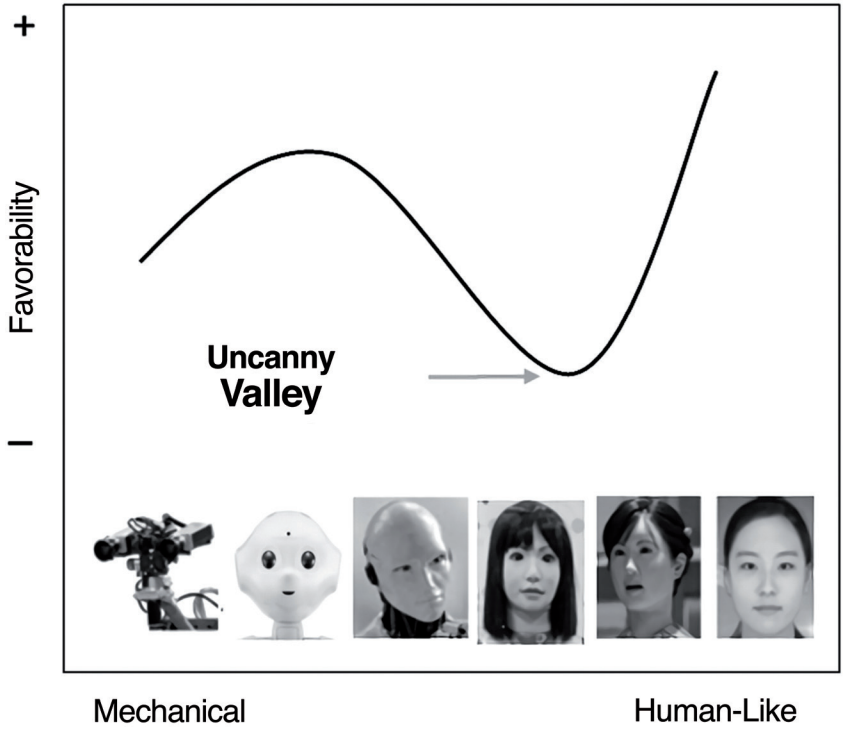
---

This chapter was originally published as Jung, Y., Cho, E., & Kim, S. (2021). Users' affective and cognitive responses to humanoid robots in different expertise service contexts. *Cyberpsychology, Behavior, and Social Networking, 24*(5), 300-306. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2019S1A3A2099973) and the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2020-0-01749) supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation).

ers, their impersonality and inability to understand informal language (e.g., slang) are reportedly regarded as significant concerns (Travelzoo, 2017). Accordingly, the diffusion of robots in the social realm increases individuals' interactions with robots and requires a profound understanding of human-robot interactions (HRIs).

The uncanny valley (UCV), which refers to the negative feeling humans derive from interactions with even highly realistic human-like robots (Mori, 1970), could be another significant cause of individuals' unfavorable feelings for robots. Although some researchers have argued that anthropomorphizing robots has a positive effect on individuals' attitude (van Doorn et al., 2017), many studies have produced inconsistent or even contrary results (Lu et al., 2019) in this regard. The UCV theory can explain such contingent findings. The theory posits the existence of a nonlinear relationship between anthropomorphism and favorability; while anthropomorphized robots basically increase favorability, their imperfect resemblance to humans can cause extremely negative feelings (Figure 1).

Figure 1 - *The uncanny valley*

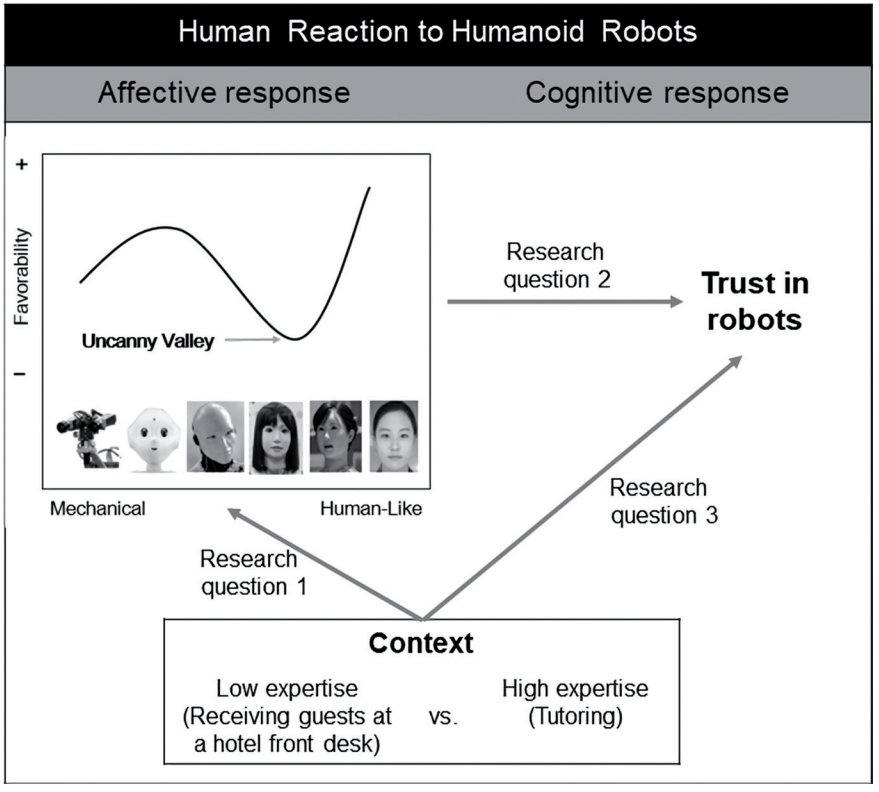


The use of robots for individual users or customers is increasingly widespread in diverse social and business contexts. In particular, robots empowered by artificial intelligence (AI) can operate in ever more diverse service contexts (Huang & Rust, 2013). This trend highlights the importance of examining UCV in specific use contexts. Users' responses to robots have reportedly been significantly influenced by the types of interactions they engage in with robots (Oyedele et al., 2007; Gaudiello et al., 2016). Although the effects of human likeness or anthropomorphism can be contingent on situational factors (Ketron & Naletelich, 2019), our knowledge of how the UCV works in different application contexts remains limited. Thus, our first research question was as follows: *Does the context of human interactions with humanoid robots affect the uncanny valley?* More specifically, this study investigated the UCV in different expertise contexts. Advances in AI technologies have resulted in the creation of humanoid robots with diverse levels of expertise and led to their deployment in diverse contexts (Gonzalez-Jimenez, 2018). As users have more interaction with diverse robots with different expertise, the expertise context becomes more significant in understanding users' responses to robots. Another reason for choosing the expertise context is that robots' expertise can affect users' attitude toward robots. The expertise of service providers is regarded as a significant factor influencing consumers' evaluation of service quality (Parasuraman et al., 1988) and such effect has been demonstrated in diverse contexts of user adoption of technology, including chatbots (Chung et al., 2020). We therefore think that the UCV (i.e., attitude toward robots) can be affected by their expertise. Additionally, expertise is a dimension of trust (McKnight et al., 2002), and thus, we expect that robots' expertise is a significant factor influencing users' overall trust in robots, which is examined as users' cognitive response in this study. Prior research also posits that the expertise of chatbots affects consumers' trust in services (Nordheim et al., 2019).

Another limitation of the UCV hypothesis is that it only examines emotional responses to humanoid robots without considering cognitive responses. Since users' interactions with humanoid robots are growing in volume and sophistication, a more comprehensive understanding of their reactions is required. Both affective and cognitive processes contribute to human decision-making and behavior (Loewenstein et al., 2001). Prior research has shown that individuals' cognitive assessments of humanoid robots are significantly related to the affective responses presented in the UCV (Mathur & Reichling, 2016). In particular, trust in robots can be an important cognitive factor in humans' interaction with robots (Gaudiello et al., 2016). Humanoid robots imitate human characteristics and behave in human-like ways. People may assume that their interactions with humanoid robots are comparable to interactions with humans; in other

words, the individual (trustor) expects the robot (trustee) to perform a particular action in his or her own interest. The unfamiliarity of interactions with humanoids, however, can provoke suspicion even regarding satisfactory interactions. Trust in technologies has typically been regarded as a central factor in their adoption (McKnight, 2005). Similarly, trust in humanoid robots plays a critical role in HRIs and may be essential in the current preliminary stage of service robot development. Recognizing that affective responses influence the formation of trust in other parties during interactions (Dunn & Schweitzer, 2005), our second and third research questions were as follows: *How is trust in humanoid robots reflected in the affective responses presented in the uncanny valley? Do the contexts of interactions with humanoid robots influence humans' trust in robots?* Given that trust could reflect context-dependent affective responses, we assumed that trust could be influenced by the context. Empirical research has also shown that environmental factors significantly impact trust in robots (Hancock et al., 2011). Figure 2 shows our research model.

Figure 2 - *Research model*



Trust in Humanoid Robots

Although UCV has been examined extensively, previous UCV studies have not considered individuals' cognitive responses to various anthropomorphized robots. In this study, we examined the relationship between UCV (emotional reaction) and trust (cognitive reaction) in robots. Trust reflects an individual's willingness to make themselves vulnerable to another party's behavior (Rousseau et al., 1998). Trust facilitates interactions among individuals and has been widely used to explain individuals' behavior in numerous computer-meditated environments (Gefen et al., 2003).

While many studies have employed *trust-in-people* to explain individuals' Internet-mediated interaction practices, some studies have assumed that information technology (IT) itself can be a trustee and examined *trust-in-technology*. Given that trust formation depends on the characteristics of counter parties (Rousseau et al., 1998), individuals may evaluate the trustworthiness of a given IT based on the attributes it manifests in an interactive context. For example, Wang and Benbasat (2005) revealed that trust is a critical factor in users' interactions with online recommendation agents. In their study, they regarded the recommendation agent software as a trustee. Similarly, people may assess humanoid robots' trustworthiness when interacting with them. People may even be more inclined to treat humanoid robots more like humans than other types of IT since humanoid robots have the characteristics of humans in terms of appearance and they engage in social interactions with other humans (Groom & Nass, 2007). In fact, trust has been regarded as a critical indicator in assessments of the quality of human-computer interactions (Lee & See, 2004). Its persuasive function in social interactions could affect individuals' acceptance of robots (Salem et al., 2015).

Hancock and colleagues (2011) conducted a meta-analysis to examine the factors that influence trust in robots by classifying them as user-related (e.g., prior experience), robot-related (e.g., robot attributes), and environmental factors (task type). They found that while robot-related and environmental factors significantly influenced trust in robots, user-related factors had little effect. Supposing that the UCV mirrors robots' traits, trust could be closely related to the UCV. Individuals' emotional states influence their decision-making (Vohs et al., 2014). When individuals make judgments, their feelings serve as critical reference points (Schwarz & Clore, 1988). The relationship between affect and decision-making also applies in trust formation. Although rational models of trust posit that trust development depends on careful and deliberate processing, trust is significantly influenced

by affect, which is derived from available cues (Lount, 2010). It has been argued that individuals with positive emotions are more likely to trust other parties (Dunn & Schweitzer, 2005). The significant impact of feeling on believing has been employed in IT contexts. Research has revealed that the affective responses presented in the UCV deeply influence individuals' assessment of humanoid robots' trustworthiness (Loewenstein et al., 2001).

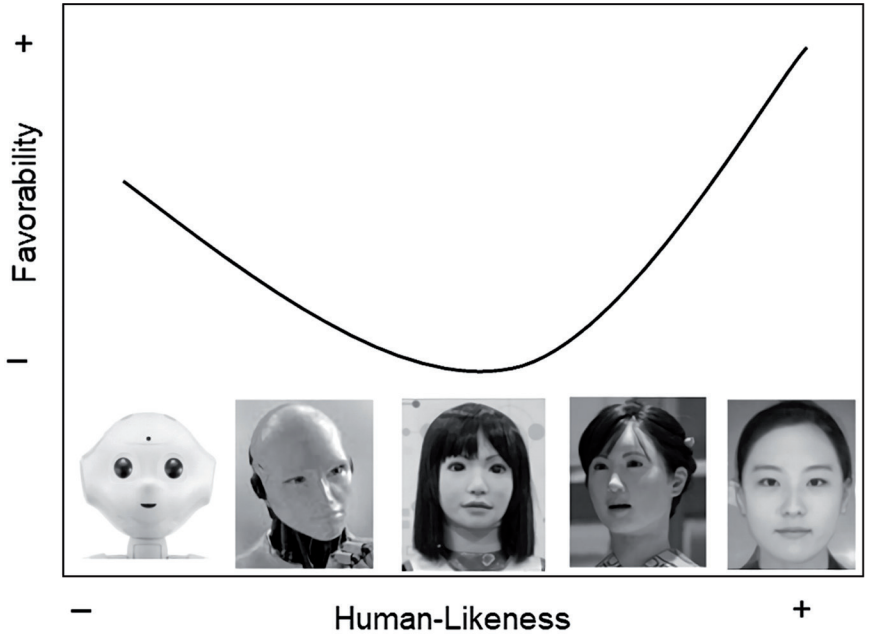
As Hancock et al. (2011) found, environmental components also influence trust in robots. Some studies have examined the contextual impact of trust in robots. Gaudiello et al. (2016) found that users have more trust in robots in functional contexts than in social contexts. Salem et al. (2015) posited that the revocability of the outcomes of tasks conducted by robots affects users' acceptance of robots' recommendations. The fact that the use of robots in various personal and social contexts has become increasingly common highlights the need for further investigation of robot trust in a wider range of contexts.

Methodology

To select pictures of humanoids to present in an experimental questionnaire, we reviewed pictures of humanoids that were used in previous relevant studies as well as photos on the Internet that we found using the keyword 'humanoid'. In the initial phase, we collected 10 humanoid pictures. Through a pilot test that asked about the human resemblance of each picture, we ultimately chose five pictures that presented a clear UCV gradation for parsimonious analyses (Figure 3). In addition, in the pilot test, we checked the expertise levels required in two task contexts: receiving guests at a hotel front desk versus tutoring. We chose these two situations because they have been widely presented as possible or actual applications of intelligent robots (Lu et al., 2019). Prior literature suggests that executing transactions with customers is regarded as a simple task for service robots, whereas educating customers is a professional job for robots (Paluch et al., 2020). Results of our pilot test confirmed that respondents perceived the different expertise levels of two contexts (t=2.910, p<0.01).

We adopted measurement questions that have been used in prior studies of the UCV (Gray & Wegner, 2012) and added questions regarding the degree of human likeness of humanoids ('Do you agree that the robot in this picture looks like a human?') and measuring favorability ('Do you feel uneasy or unfriendly when looking at the robot in this picture?'). We also developed a question to measure overall trust in robots: 'Do you believe that the robot in this picture would be

Figure 3 - *Humanoid pictures in an experimental questionnaire*



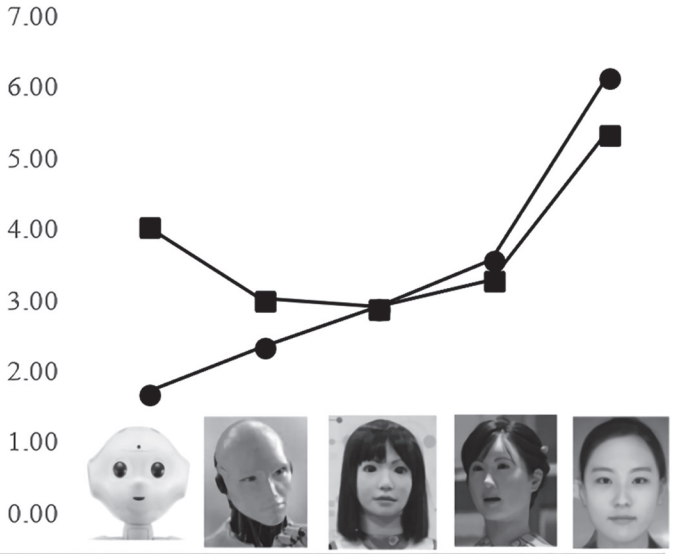
trustworthy as a hotel reception staff (or tutor)?' All items were measured using 7-point Likert scales ranging from 'strongly disagree' to 'strongly agree'.

We conducted a between-subject experimental survey in which participants provided answers to questions about their affective responses (favorability) and cognitive responses (trust) to five different types of humanoid robot photos in one of two situations (receiving guests at a hotel front desk vs. tutoring). We collected data from the panel members of an online research firm in South Korea and administered web-based online surveys over a 1-week period in March 2019. We gathered data separately for the two study contexts. After we eliminated invalid responses, we included a final sample of 505 participants in the analysis (251 in the hotel reception context and 254 in the tutoring context). The mean age of the participants was 39.3 years. In the first stage of the survey, participants assessed the human likeness of five humanoid pictures. Next, they read a short scenario describing the context and provided their responses regarding favorability and trust for each of five humanoid pictures, which were randomly presented. Finally, participants answered demographic questions.

Results

The results of a paired difference test confirmed a human-likeness hierarchy for the five humanoid pictures, whose favorability showed a U-curve, evidence of the UCV effect, as did the results of the pilot test (Figure 4).

Figure 4 - *Human likeness and favorability of five humanoid pictures*



Human-likeness	1.71	2.36	2.93	3.58	6.16
Favorability	4.06	3.03	2.92	3.30	5.37

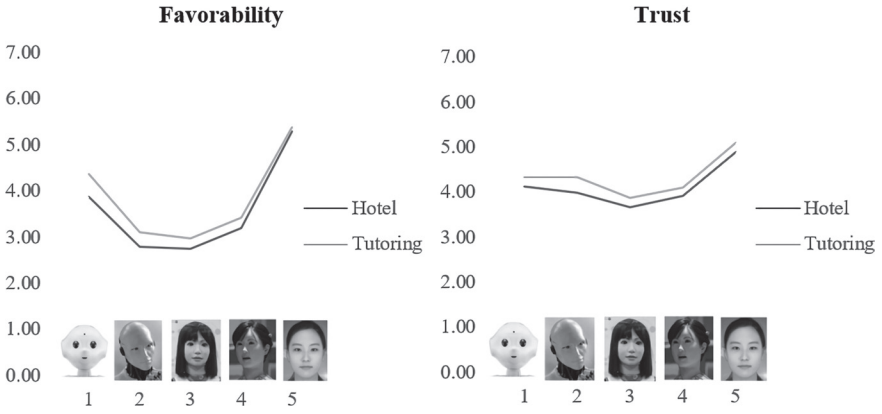
We conducted analysis of variance (ANOVA) to examine differences by context. The data achieved ANOVAs assumption of homogeneity of variance, and the results showed significant contextual differences in favorability and trust for all stages except for favorability in the last stage (Table 1 and Figure 5). Participants evaluated both favorability and trust more positively in the tutoring context than in the hotel reception context, but the *F*-values (mean differences) revealed a variation in this result. The contextual effect on participants' appraisals was more powerful for favorability in stages I, II, and III. We also conducted a regression analysis to examine the influence of favorability on trust. In the hotel reception context, the adjusted R^2 was 0.308 and the coefficient of favorability was 0.557 (t=10.595, p=0.000); in the tutoring context, the adjust-

Table 1 - *Analysis of variance results for favorability and trust by context*

	M	SD	*Homogeneity test*		*Difference of means*	
			Levene's statistic	p	F	p
Favorability						
Stage I						
Hotel reception	3.87	1.28	0.39	0.55	14.97	0.00
Tutoring	4.36	1.33				
Stage II						
Hotel reception	2.80	1.05	0.45	0.50	9.99	0.00
Tutoring	3.11	1.14				
Stage III						
Hotel reception	2.75	1.04	0.92	0.34	5.15	0.02
Tutoring	2.97	1.13				
Stage IV						
Hotel reception	3.20	1.10	2.59	0.11	4.99	0.03
Tutoring	3.43	1.23				
Stage V						
Hotel reception	5.29	1.19	2.34	0.13	0.87	0.35
Tutoring	5.39	1.33				
Trust						
Stage I						
Hotel reception	4.14	1.15	0.43	0.51	4.03	0.04
Tutoring	4.35	1.10				
Stage II						
Hotel reception	4.00	1.21	0.74	0.39	8.62	0.00
Tutoring	4.33	1.27				
Stage III						
Hotel reception	3.68	1.13	0.98	0.32	4.14	0.04
Tutoring	3.89	1.20				
Stage IV						
Hotel reception	3.93	1.11	0.87	0.35	4.70	0.03
Tutoring	4.12	1.18				
Stage V						
Hotel reception	4.89	1.22	0.28	0.60	4.71	0.03
Tutoring	5.11	1.09				

ed R^2 was 0.310 and the coefficient of favorability was 0.559 (t=10.708, p=0.000). These results show that favorability had a significant impact on trust.

Figure 5 - *Comparison of favorability and trust by context*



Discussion

This study contributes to service robot research by elucidating the effects of robots' anthropomorphism. Human-like robots or humanoids have been broadly utilized in service areas, and thus, anthropomorphism has been regarded as a significant factor that influences consumers' attitudes toward humanoid robots. Previous studies have produced inconsistent findings regarding the effects of anthropomorphized robots. Many studies have demonstrated that robot anthropomorphism generates warmth (Van Doorn et al., 2017) or adoption intentions (Tussyadiah & Park, 2018), whereas others have revealed anthropomorphism's contingent (Mende et al., 2019) or negative (Lu et al., 2019) effects. This study explicates anthropomorphism's effects in service fields by employing the UCV theory.

Furthermore, this study provides a better account of these effects by showing perceptual differences regarding anthropomorphism in different service contexts. Applying the UCV perspective, this study examined how peoples' affective and cognitive responses to humanoid robots differed in two service contexts, hotel reception and tutoring. As far as the power of the UCV is concerned, our analyses revealed ambivalent results. On the one hand, we found that favorability had a significant effect on trust, implying that affective appraisals of humanoids play a role in

initial impressions, which are the foundation for further cognitive evaluations. This finding, therefore, suggests that the UCV describes the principal reaction to humanoids, and its influence is not limited to affective responses but applicable to cognitive responses to humanoids as well. On the other hand, the depth of the UCV (i.e., differences in favorability across the stages) decreased trust. This result implies that the impact of the UCV was attenuated by trust in robots. Because interactional evaluations (e.g., trust) can form primary attitudes toward robots in actual real-world settings, the impact of the UCV might be weakened in real contexts. Conclusively, although the UCV indicating affective evaluation of robots is influential in people's attitude toward them, its impact might be declined in actual contexts. Our findings suggest that researchers should examine individuals' affective and cognitive responses to robots in actual application environments while also scrutinizing how such responses lead to outcome behaviors, such as continuous use, service satisfaction, and compliance with robots (Lee & Liang, 2016).

Our analysis also revealed that affective and cognitive responses were more positive for the high-expertise humanoid (tutoring) than for the low-expertise humanoid (hotel reception) in all stages of the UCV except for the last stage, where the humanoid's face is the same as a human's face. This finding suggests that when people form impressions of humanoids conducting certain tasks for them, their assessments differ based on the task type. More specifically, people's attitudes are less influenced by humanoids' peripheral cues (e.g., appearances) in tasks requiring higher levels of expertise. This result can be explained by the elaboration likelihood theory, which proposes that humans process stimuli via two distinct routs (i.e., a central route and a peripheral route) (Petty & Cacioppo, 1986). When elaborating on a message, the purposeful and conscious processing of an argument (stimulus) based on an individual's ability occurs via a central route while a heuristic processing based on general impressions occurs via a peripheral route (Petty & Cacioppo, 1986). In our context, when a humanoid's task is complicated or knowledge-intensive, people's dominant attention on its ability to successfully complete a task mitigates the influence of the humanoid's appearance on their reaction to it. This result implies that people may differently evaluate robots based on the robots' level of expertise. Accordingly, researchers need to examine the UCV phenomenon in diverse real-world contexts (e.g., static robots vs. movable robots, intraorganizational contexts vs. consumer service contexts). Managers should also take service type into consideration when developing service robots; in particular, managers trying to use simple service robots must be careful about the UCV. Managers might be better off using modest human-like robots (first stage example in our experiment) to offer simple services to

customers rather than incomplete human-like ones (second, third, and fourth stage examples in our experiment).

One limitation of this study is the use of a single item for measuring each factor. The study employed it to mitigate respondents' fatigue of a long survey process. Instead, we asked fundamental and straightforward questions (e.g., 'Do you agree that the robot in this picture looks like a human?' for measuring human likeness) to reduce the weakness of using a single-question scale. However, there are well-developed measuring scales of anthropomorphism or human likeness (Bartneck et al., 2009), trust (Madsen & Gregor, 200), and favorability. Accordingly, future research is recommended to employ multiscale measurement to achieve rigorousness.

As robots are equipped with AI and HRI becomes prevalent in the social and commercial lives, consequences of such interaction can be more complicated and unpredictable. As revealing the dynamics of humans' responses to robots, this study contributes to understanding of future HRI. Researchers also envisage that advances in AI will make robots more intelligent and even have a sense of self within the next few decades (Baum et al., 2011). Such evolution of robots implies that they will be able to fully acquire knowledge and make decisions, and ultimately be indistinguishable from humans (Gonzalez-Jimenez, 2018). Accordingly, future research needs to examine the UCV in human interaction with more intelligent or emotional robots beyond the current focus on robots' tangible attributes such as appearance or voice (Jung & Cho, 2018).

References

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics, 1*(1), 71-81.

Baum, S. D., Goertzel, B., & Goertzel, T. G. (2011). How long until human-level AI? Results from an expert assessment. *Technological Forecasting and Social Change, 78*(1), 185-195.

Benbasat, I., & Wang, W. (2005). Trust in and adoption of online recommendation agents. *Journal of the Association for Information Systems, 6*(3), 4.

Chung, M., Ko, E., Joung, H., & Kim, S. J. (2020). Chatbot e-service and customer satisfaction regarding luxury brands. *Journal of Business Research, 117*, 587-595.

Dunn, J. R., & Schweitzer, M. E. (2005). Feeling and believing: The influence of emotion on trust. *Journal of personality and social psychology, 88*(5), 736.

Gaudiello, I., Zibetti, E., Lefort, S., Chetouani, M., & Ivaldi, S. (2016). Trust as in-

dicator of robot functional and social acceptance. An experimental study on user conformation to iCub answers. *Computers in Human Behavior, 61*, 633-655.

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS quarterly, 27*, 51-90.

Gonzalez-Jimenez, H. (2018). Taking the fiction out of science fiction: (Self-aware) robots and what they mean for society, retailers and marketers. *Futures, 98*, 49-56.

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition, 125*(1), 125-130.

Groom, V., & Nass, C. (2007). Can robots be teammates?: Benchmarks in human-robot teams. *Interaction studies, 8*(3), 483-500.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human factors, 53*(5), 517-527.

Huang, M. H., & Rust, R. T. (2013). IT-related service: A multidisciplinary perspective. *Journal of Service Research, 16*(3), 251-258.

Jung, Y., & Cho, E. (2018). Context-specific affective and cognitive responses to humanoid robots. In *Proceedings of the 22nd International Telecommunications Society (ITS) Biennial Conference*. ITS.

Ketron, S., & Naletelich, K. (2019). Victim or beggar? Anthropomorphic messengers and the savior effect in consumer sustainability behavior. *Journal of Business Research, 96*, 73-84.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human factors, 46*(1), 50-80.

Lee, S. A., & Liang, Y. (2016). The role of reciprocity in verbally persuasive robots. *Cyberpsychology, Behavior, and Social Networking, 19*(8), 524-527.

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological bulletin, 127*(2), 267.

Lount Jr, R. B. (2010). The impact of positive mood on trust in interpersonal and intergroup interactions. *Journal of personality and social psychology, 98*(3), 420.

Lu, L., Cai, R., & Gursoy, D. (2019). Developing and validating a service robot integration willingness scale. *International Journal of Hospitality Management, 80*, 36-51.

Madsen, M., & Gregor, S. (2000). Measuring human-computer trust. In *11th australasian conference on information systems: Vol. 53* (pp. 6-8). Australasian Association for Information Systems.

Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. *Cognition, 146*, 22-32.

McKnight, D. H. (2005). Trust in information technology. *The Blackwell encyclopedia of management, 7*, 329-331.

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating

trust measures for e-commerce: An integrative typology. *Information systems research, 13*(3), 334-359.

Mende, M., Scott, M. L., van Doorn, J., Grewal, D., & Shanks, I. (2019). Service robots rising: How humanoid robots influence service experiences and elicit compensatory consumer responses. *Journal of Marketing Research, 56*(4), 535-556.

Mori, M. (1970). Bukimi no tani [The uncanny valley]. *Energy, 7*, 33-35.

Nichols, G. (2018, January 22). Robot fired from grocery store for utter incompetence. ZDNet. https://www.zdnet.com/article/robot-fired-from-grocery-store-for-utter-incompetence/.

Nordheim, C. B., Følstad, A., & Bjørkli, C. A. (2019). An initial model of trust in chatbots for customer service - findings from a questionnaire study. *Interacting with Computers, 31*(3), 317-335.

Oyedele, A., Hong, S., & Minor, M. S. (2007). Contextual factors in the appearance of consumer robots: exploratory assessment of perceived anxiety toward humanlike consumer robots. *CyberPsychology & Behavior, 10*(5), 624-632.

Paluch, S., Wirtz, J., & Kunz, W. H. (2020). Service robots and the future of service. In M. Bruhn, M. Kirchgeorg & C. Burmann (Eds.), *Marketing Weiterdenken - Zukunftspfade für eine marktorientierte Unternehmensführung (2nd Ed.)* (pp. 423-435). Springer.

Parasuraman, A., Zeithaml, V. A., & Berry, L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing, 64*(1), 12-40.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. Springer-Verlag.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of management review, 23*(3), 393-404.

Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015). Would you trust a (faulty) robot? Effects of error, task type and personality on human-robot cooperation and trust. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 1-8). IEEE.

Schwarz, N., & Clore, G. L. (1988). How do I feel about it? Informative functions of affective states. In K. Fiedler & J. Forgas (Eds.), *Affect, Cognition, and Social Behavior* (pp. 44-62). Hogrefe.

Travelzoo. (2017, December 6). The travel Industry is relying on robots more and more—and most of us are fine with that. *Huffington Post*. https://www.huffpost.com/entry/the-travel-industry-is-re_b_9421890.

Tussyadiah, I. P., & Park, S. (2018). Consumer evaluation of hotel service robots. In B. Stangl & J. Pesonen (Eds.), *Information and Communication Technologies in Tourism 2018* (pp. 308-320). Springer.

Van Doorn, J., Mende, M., Noble, S. M., Hulland, J., Ostrom, A. L., Grewal, D., & Petersen, J. A. (2017). Domo arigato Mr. Roboto: Emergence of automated social

presence in organizational frontlines and customers' service experiences. *Journal of service research, 20*(1), 43-58.

Vohs, K. D., Baumeister, R. F., Schmeichel, B. J., Twenge, J. M., Nelson, N. M., & Tice, D. M. (2014). Making choices impairs subsequent self-control: A limited-resource account of decision making, self-regulation, and active initiative. *Journal of personality and social psychology, 94*, 883-898.

2. Emerging Adults' Expectations about the Next Generation of Robots

Exploring Robotic Needs Through a Latent Profile Analysis

F. Manzi, A. Sorgente, D. Massaro, D. Villani, D. Di Lernia, C. Malighetti, A. Gaggioli, D. Rossignoli, G. Sandini, A. Sciutti, F. Rea, M.A. Maggioni, A. Marchetti, G. Riva

ABSTRACT

The investigation of emerging adults' expectations of development of the next generation of robots is a fundamental challenge to narrow the gap between expectations and real technological advances, which can potentially impact the effectiveness of future interactions between humans and robots. Furthermore, the literature highlights the important role played by negative attitudes toward robots in setting people's expectations. To better explore these expectations, we administered the Scale for Robotic Needs and performed a latent profile analysis to describe different expectation profiles about the development of future robots. The profiles identified through this methodology can be placed along a continuum of robots' humanization: from a group that desires mainly the technical features to a group that imagines a humanized robot in the future. Finally, the analysis of emerging adults' knowledge about robots and their negative attitudes toward robots allowed us to understand how these affect their expectations.

Introduction

The significant advances in robotic research suggest a future in which robots will play an increasingly important role in different contexts of human life. In recent years, we have witnessed the introduction of robots in different contexts, such as home and work (Sung et al., 2010; Tapus et al., 2007), and in sensitive domains of our society, such as in ed-

---

This chapter was originally published as Manzi, F., Sorgente, A., Massaro, D., Villani, D., Di Lernia, D., Malighetti, C., Gaggioli, A., Rossignoli, D., Sandini, G., Sciutti, A., Rea, F., Maggioni, M.A., Marchetti, A., & Riva, G. (2021). Emerging adults' expectations about the next generation of robots: Exploring robotic needs through a latent profile analysis. *Cyberpsychology, Behavior, and Social Networking, 24*(5), 315-323. No competing financial interests exist. This research was funded by Università Cattolica del Sacro Cuore (D3.2–2018–Human-Robot Confluence project). Supplementary Data.

ucational contexts or intervention and rehabilitation practices (Belpaeme et al., 2018; Bemelmans et al., 2012; Marchetti et al., 2020; Serino et al., 2018; Riva et al., 2019). The use of robots in these contexts is particularly important for those parts of the population that will encounter the need to use them in their daily routines and especially at work. One of these groups is certainly that of emerging adults (people aged 18-29 years) (Arnett, 2000): among the various challenges of their development, they also have to deal with the world of work where they will increasingly have to integrate robots into their work routines.

To ensure the effectiveness of robots in social contexts, they should become technically efficient and credible social partners capable of sustaining long-term interactions with humans (Leite et al., 2013; Sciutti et al., 2018; Vannucci et al., 2019). However, there is still a long way to go in terms of technology to achieve these goals. Despite these technical limitations, several studies have shown that humans attribute human-like qualities to robots even in their current form (Di Dio et al., 2020a,b,c; Manzi et al., 2020a,b,c; 2021; Marchetti et al., 2018): this means that people have expectations of robots in terms of performance and interaction possibilities. However, a gap between people's expectations and actual technological advancements of robots could result in less effective human-robot interactions in the future (Kwon et al., 2016; Sciutti et al., 2018). More specifically, understanding people's expectations about future robots is essential to guide researchers in the development of existing and new robots.

The issue of people's expectations regarding the technical and social skills to be implemented in future robots has not yet been widely studied. Furthermore, there is no theoretical model in the literature on robots that explains the variables that influence people's expectations for development of future robots. Although there is no specific theoretical model for robots, the literature identified an association between attitudes toward robots and expectations toward an ideal robot (Kuhnert Ragni & Lindner, 2017). This association between attitudes and expectations was also found in psychology in human interactions (Thompson & Sunol, 1995).

In addition, psychological literature reports the effects of other variables on expectations, grouped into two macrocategories: personal and social variables (Thompson & Sunol, 1995). In particular, knowledge and attitudes belong to the personal domain, while sociodemographic characteristics belong to the social domain. Based on this lack of knowledge in robotics and the importance of other factors influencing people's expectations, the present study intends to explore the influence of sociodemographic characteristics, knowledge about robots, and attitudes toward robots on people's expectations of future robots.

An important point to consider is related to the type of people's work; in fact, the Eurobarometer (2012) showed (in the European population) that the concern about robots as a dangerous technology (i.e., a technology that can steal jobs) is high among manual workers, while it is lower for managers. An aspect that emerged from the Eurobarometer (2012) shows that people with a higher level of education tend to rate robots (i.e., a technology to help humans) more positively than those with a lower level of education.

In addition, European data show a positive trend in the general interest in technological developments, although the percentage of people who have interacted with a robot is quite low (around 12 percent) (Eurobarometer, 2012). European citizens also accept the idea of using a robot in the manufacturing context more than their domestic use. Moreover, people who would gladly use it at home are more inclined to employ it for household purposes and only a small percentage would use it for company (Eurobarometer, 2012).

Another factor that can affect the expectations of emerging adults is gender. Indeed, some studies suggested that gender affects how robots are perceived in terms of possible impact on society (Lin et al., 2012; Loffredo & Tavakkoli, 2016; Nomura et al., 2006). For example, males perceive robots as more useful technology for society than females (Loffredo & Tavakkoli, 2016). However, these studies also demonstrate the absence of differences compared with other aspects such as the need to control robots, for which both genders score high (Loffredo & Tavakkoli, 2016). Therefore, although there are data showing a gender difference, the debate is still open.

Finally, many studies have analyzed people's negative attitudes toward robots and how these can affect interactions between humans and robots and people's propensity to include robots in sensitive contexts such as clinical or educational context (Bartneck et al., 2007; Nomura et al., 2006). These studies delineate a complex scenario with respect to the link between attitudes toward robots and desire to interact with them. A recent study by Billing et al. (2019) showed a correlation between negative attitudes toward robots and expectations of using them in a clinical context, in particular showing that fewer negative attitudes toward robots positively affect people's expectations. However, since there is no extensive literature yet on this topic, we believe that our study could shed light on a more complex scenario concerning people's expectations and negative attitudes toward robots.

In this sense, the present study aims to identify profiles describing the different emerging adults' expectations about the development of technical and interactive aspects of future robots and to examine the in-

fluence of sociodemographic characteristics, knowledge about robots, and negative attitudes toward robots on these expectations.

Materials and Methods

Participants

Participants were 344 Italian emerging adults (57 percent female) aged 18-29 years (M=24.94; standard deviation [SD]=3.07). Around half of the sample (54.7 percent) did not have any academic degree (their highest education level was the high school diploma or lower). Emerging adults having a bachelor's/master's degree or a higher level of education, such as a PhD, were instead 45.3 percent of the sample. One-third of the sample (35.9 percent) comprised workers, while others were students (51.4 percent) or unemployed (12.7 percent). Finally, 7.3 percent of the sample was married or cohabiting.

This sample was collected on two different occasions: 41.6 percent of the participants filled in the survey in 2019 (as this data collection was needed to perform a pilot study; see Supplementary Data)[1], while the remaining 58.4 percent filled in the survey in 2020. In both cases, emerging adults signed an informed consent and were treated in accordance with the Declaration of Helsinki.

Instruments

The description of administered instruments is reported in Table 1.

Data analysis

Latent profile analysis: identifying different profiles of expectations about future robots. To identify the groups (i.e., profiles) that best describe the heterogeneity within the current sample with respect to the different expectations people have about future robots, we performed a latent profile analysis (LPA) using Mplus software, including the four-factor scores of the Scale for Robotic Needs (SRN) (technical features, social-emotional resonance, agency, and human life) as observed indicators. We examined fit indices of measurement models, beginning with one class and adding classes incrementally. As suggested by

[1] https://www.liebertpub.com/doi/suppl/10.1089/cyber.2020.0161/suppl_file/Supp_Data.zip.

Table 1 - *Description of administered instruments*

Scale	*Construct*	*Dimensions*	*Number of items*	*Scale response*	*Omega*
Scale for Robotic Needs (SRN)[a]	Expectations about future robots	Technical features	3	Five-point scale (1 = not at all; 5 = very much)	0.56
		Social-emotional resonance	3		0.88
		Agency	3		0.59
		Human life	8		0.94
Negative Attitudes Toward Robots Scale (NARS)	Negative attitudes toward…	…situations and interactions with robots	6	Second-point scale (1 = strongly disagree; 7 = strongly agree)	0.75
		…social influence of robots	5		0.75
		…emotions in interactions with robots	3		0.76
Knowledge about robots	General knowledge	Overall, how is your knowledge about robots?	1	Five-point scale (1 = no knowledge; 5 = professional knowledge)	—
	Definition	According to you, is the robot a dangerous technology for humans?	1	Yes/No	—
		According to you, is the robot a technology that can help humans?	1	Yes/No	—
		According to you, is the robot a technology that can keep humans company?	1	Yes/No	—
	Applied domains	According to you, can robots be applied to a domestic domain?	1	Yes/No	—
		According to you, can robots be applied to a clinic domain?	1	Yes/No	—
		According to you, can robots be applied to a business domain?	1	Yes/No	—

Note: [a]The expected four-factor structure was tested on the current sample, showing good fit indices [$\chi^2(113) = 197.975$; $p < 0.001$; RMSEA (90 percent CI) = 0.061 (0.047–0.075); $p = 0.097$; CFI = 0.944; SRMR = 0.061]. Information about this instrument development and evidence of its score validity are reported in the Supplementary Data[2]. See Appendix A1 for exact item wording. CI, confidence interval; RMSEA, root-mean-squared error of approximation; CFI, comparative fit index; SRMR, standardized root mean of the residual.

[2] https://www.liebertpub.com/doi/suppl/10.1089/cyber.2020.0161/suppl_file/Supp_Data.zip.

Sorgente and colleagues (2019) we used different measures of relative model fit to make decisions about the best class solution.

Specifically, we used two information criteria: the Akaike (1974) information criterion (AIC) and Bayesian information criterion (BIC; Schwarz, 1978), together with another descriptive measure of relative model fit: the approximate Bayes factor (BF), which compares two models at a time (k and $k+1$ model, where k is the number of classes) and the best model is the most parsimonious k-class model with BF >3. Furthermore, for each LPA solution, we evaluated the Lo-Mendell-Rubin likelihood ratio test (adjusted LMR-LRT; Nagin, 1999) that compares a $(k-1)$-class model with a k-class model; if it is not significant, the k-class model is as good as the $(k-1)$-class model, so the $(k-1)$-class model is preferred according to the parsimony criterion. Finally, we compared models evaluating their level of entropy, where values closer to 1 indicate better classification of cases (Masyn, 2013).

Once the best solution(s) is selected, the precision of this solution in assigning individuals to a class can be evaluated by (a) comparing the class proportion (π_k) with the modal class assignment proportion ($mcaP_k$) and (b) estimating the average posterior probability ($avePP_k$); as well as (c) estimating the odds of correct classification (OCC_k; Masyn, 2013). The precision of the latent class solution is good when the $mcaP_k$ for each class is included in the 95 percent confidence interval of the π_k. Furthermore, to suggest well-separated classes, the $avePP_k$ should be 0.70 or higher and the OCC_k should be above 5.

Chi-square tests and regressions: identifying predictors of the profiles' membership. Once we identified the different profiles describing the diverse expectations that people have about future robots, we verified, through the chi-square test or logistic regression, if sociodemographic variables (age, gender, educational level, and occupational status), knowledge about robots (general knowledge, definition, and applied domains), and negative attitudes toward robots (three factors of the Negative Attitudes Toward Robots Scale [NARS]) could predict the expectation profile to which people belong.

Results

Descriptive statistics

Descriptive statistics referred to the variables investigated in the current study (expectation about future robots, negative attitudes toward robots, and knowledge people have about robots) are reported in Table 2.

Table 2 - *Descriptive statistics for variables measuring expectations about future robots, negative attitudes toward robots, and knowledge (general, definition, and applied domains) people have about robots (N=344)*

		M (*SD*)
Expectations about future robots	Technical features	4.01 (0.57)
	Social-emotional resonance	3.14 (1.03)
	Agency	3.76 (0.66)
	Human life	1.80 (0.95)
Negative attitudes toward…	…situations and interactions with robots	2.62 (0.94)
	…social influence of robots	3.79 (1.20)
	…emotions in interactions with robots	4.10 (1.24)
General knowledge about robots	Overall, how is your knowledge about robots?	2.77 (0.69)
		Number of 'yes' (percent)
Definition	According to you, is the robot a dangerous technology for humans?	8 (3.7)
	According to you, is the robot a technology that can help humans?	200 (93.0)
	According to you, is the robot a technology that can keep humans company?	25 (11.6)
Applied domains	According to you, can robots be applied to a domestic domain?	179 (83.3)
	According to you, can robots be applied to a clinic domain?	179 (83.3)
	According to you, can robots be applied to a business domain?	195 (90.7)

Note: SD, standard deviation.

Latent profile analysis

The four-factor scores measuring, respectively, future robots' technical features, social-emotional resonance, agency, and human life were extracted from the confirmatory factor analysis model of the SRN and used as observed indicators of the LPA. As shown in Table 3, the five-class solution was the preferred one. Despite AIC and BIC being poorly informative as they improved when the number of classes increased, BF, the adjusted LMR-LRT, and the level of entropy suggested retaining the five-class solution.

Table 3 - *Relative model fit indices for the seven latent profile models*

Model	*AIC*	*BIC*	*BF*	*LMR-LRT*	*E*	*Class size*
1-Class	3224.26	3278.02	0.00			344
2-Class	3093.77	3166.74	0.00	<0.001	0.925	299-45
3-Class	3028.59	3120.77	0.00	0.006	0.892	87-225-32
4-Class	2963.89	3075.27	0.00	0.016	0.938	200-81-39-24
5-Class	2922.47	3053.05	**6.56**	**0.012**	**0.953**	198-39-76-18-13
6-Class	2907.03	3056.81	0.13	0.121	0.947	32-195-13-12-71-21
7-Class	**2883.68**	**3052.67**	0.00	0.250	0.936	20-32-60-167-41-11-13

Note: Information criterion values in bold indicate the best model fit.
AIC, Akaike information criterion; BIC, Bayesian information criterion; BF, Bayes factor; LMR-LRT, Lo-Mendell-Rubin likelihood ratio test; E, entropy.

Consequently, the five-class solution was investigated through the classification diagnostics. As reported in Table 4, this solution largely satisfied the classification-diagnostic criteria, indicating that the five identified profiles were well differentiated from each other.

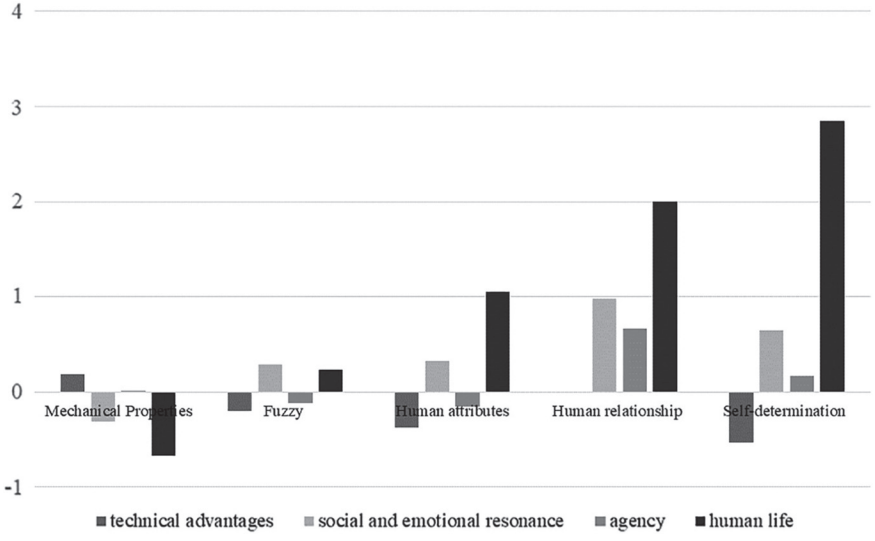
Table 4 - *Classification diagnostics for the five-class model*

Class k	*N*	*CP o* π_k *(95 percent CI)*	$mcaP_k$	$avePP_k$	OCC_k
Class 1	198	0.575 (0.520-0.632)	0.575	0.988	60.86
Class 2	76	0.110 (0.076-0.152)	0.110	0.928	104.28
Class 3	39	0.054 (0.029-0.082)	0.054	0.979	1715.06
Class 4	18	0.037 (0.018-0.061)	0.037	0.971	871.46
Class 5	13	0.223 (0.176-0.277)	0.222	0.949	330.66

Note: π_k, class proportion; $mcaP_k$, modal class assignment proportion; $avePP_k$, average posterior probability; OCC_k, odds of correct classification.

The five groups obtained (Figure 1), representing five different types of expectations that people have toward robots of the future, were named as follows: mechanical properties (N=198), fuzzy (N=76), human attributes (N=39), human relationship (N=18), and self-determination (N=13). Emerging adults of the first group scored lower on the three factors that concern interaction abilities (social-emotional resonance, agency, and human life) than average people (i.e., zero scores). The second group has been named fuzzy because people of this group scored all the four factors, giving scores that coincide with the mean score of the entire sample (i.e., their scores are included in the range of ±0.5 standard deviation from the sample mean).

Figure 1 - *The five profiles of emerging adults' expectations toward the next generation of robots obtained through administration of the Scale for Robotic Needs*



The remaining three profiles, instead, represent emerging adults who expect robots to be more similar to humans compared with other groups (i.e., the human life factor score is higher than the sample mean). These three profiles are different from each other for the emphasis given to the human life expectation and for the way in which these expectations are combined with other dimensions. In particular, the third profile (named human attributes) describes people who desire to give human attributes to the future robots, but this expectation is restrained as (a) it is lower than in the two following profiles and (b) it is combined with a

medium level of expectations about the empowerment of other abilities (technical features, social-emotional resonance, and agency).

Furthermore, the members of the fourth profile (named human relationship), other than desire empowered human-like abilities for future robots (the human life factor score is higher than the mean), they also expect that future robots will have empowered social-emotional resonance and agency skills. Finally, the mean level of the human life score of the self-determination group is three SDs above the sample mean, suggesting that they responded, 5 (=very much) to most (if not all) items measuring human life; these emerging adults are the ones who most want future robots to have human skills, such as owning money, having sex, and playing music.

Chi-square tests and regressions

A series of chi-square tests and multinomial logistic regression models were performed to verify if sociodemographic variables, levels of knowledge about robots, and negative attitudes toward robots were able to predict the expectation profile to which individuals belong.

Regarding sociodemographic variables, age [LRT: $\chi^2(4) = 4.50$; $p = 0.343$], gender [$\chi^2(4) = 6–75$; $p = 0.150$], educational level [$\chi^2(4) = 1.844$; $p = 0.764$], and occupational status [$\chi^2(8) = 8.51$; $p = 0.385$] did not predict the class membership.

Regarding the level of knowledge about robots, we first verified if the self-reported general knowledge about robots was related to the class membership, finding a nonsignificant relationship [LRT: $\chi^2(4) = 0.923$; $p = 0.921$]. Then, we investigated if the definitions of a robot endorsed by the individual affected the profile membership. We found that emerging adults belonging to the profile of mechanical properties defined a robot as a technology that can help humans [$\chi^2(4) = 10.48$; $p = 0.033$] more often than expected, while they consider the robot as a technology that can keep humans company less often than expected [$\chi^2(4) = 12.41$; $p = 0.015$]. Instead, people defining the robot as a dangerous technology for humans [$\chi^2(4) = 1.78$; $p = 0.776$] were equally distributed across the five classes. Finally, we found that the class membership did not depend on the contexts in which people think that robots can be useful: domestic [$\chi^2(4) = 2.04$; $p = 0.729$], clinic [$\chi^2(4) = 1.48$; $p = 0.831$], and business [$\chi^2(4) = 3.52$; $p = 0.475$] domains.

Regarding the negative attitudes toward robots, we performed a multinomial logistic regression in which the three factors of the NARS were included as predictors of the profiles. Results suggested that the whole regression model significantly predicted the profile member-

ship [LRT: $\chi^2(12) = 65.53$; $p < 0.001$]. Specifically, the negative attitudes toward situations and interactions with robots [$\chi^2(4) = 32.23$; $p < 0.001$] and toward emotions in interactions with robots [$\chi^2(4) = 30.67$; $p < 0.001$] significantly predicted the expectations that people have about future robots, while the negative attitudes toward the social influence of robots [$\chi^2(4) = 8.92$; $p = 0.063$] were not related to the expectations.

Table 5 summarizes the results of the regression, showing that people having high levels of negative attitudes toward situations and interactions with robots are, respectively, 0.15, 0.30, 0.32, and 0.25 times, less likely to be members of the groups, mechanical properties, fuzzy, human attributes, and human relationship, rather than the self-determination group. Instead, people having high levels of negative attitudes toward emotions in interactions with robots are, respectively, 4.45, 2.93, and 3.17 times, more likely to be members of the groups, mechanical properties, fuzzy, and human attributes, rather than the self-determination group. Instead, members of the self-determination and human relationship groups do not differ in their level of negative attitudes toward emotions in interactions with robots ($p = 0.08$).

Discussion

The current study aimed to identify specific profiles representing characteristics that should be implemented in future robots to meet emerging adults' expectations. Furthermore, the present study analyzed the effects of sociodemographic characteristics, knowledge about robots, and negative attitudes toward robots on emerging adults' expectations.

We identified five profiles of expectations that can be placed along a continuum of humanization of robots and whose ends are those who consider robots as pure technological tools at the service of humans (i.e., mechanical properties) and those who expect robots to be part of our society in the near future (i.e., self-determination).

The group mechanical properties is generally not interested in development of a robot that can 'live' as a human and, consequently, expects only a higher mechanical efficiency of robots. This peculiarity emerges when the group, mechanical properties, defines robots as a technology that can help humans. Furthermore, it is important to highlight that the same group describes robots less as a technology that can keep humans company. Therefore, there is consistency between the expectations of members of the group for mechanical properties and their definition of robots: if a person thinks that robots are tools designed to help humans (and not to interact socially), then also her/his expectations are orient-

Table 5 - *Summary of multinomial logistic regression analysis for factors of the negative attitude toward robots scale predicting membership in groups of expectations about robots of the future*

Group[a]	*Variable*	*B*	*SD*	*Wald*[b]	*p*	*OR*	*95 percent CI*
Mechanical properties	Constant	−1.06	1.07	0.99	0.32		
	NA toward situations and interactions with robots	−1.86	0.42	19.67	<0.001	0.15	0.07-0.35
	NA toward social influence of robots	0.41	0.37	1.23	0.27	1.50	0.73-3.09
	NA toward emotions in interactions with robots	1.49	0.36	16.73	<0.001	4.45	2.18-9.11
Fuzzy	Constant	1.20	1.08	1.24	0.27		
	NA toward situations and interactions with robots	−1.19	0.42	8.06	0.005	0.30	0.13-0.69
	NA toward social influence of robots	0.11	0.37	0.08	0.78	1.11	0.54-2.30
	NA toward emotions in interactions with robots	1.08	0.37	8.63	0.003	2.93	1.43-6.01
Human attributes	Constant	0.93	1.15	0.65	0.42		
	NA toward situations and interactions with robots	−1.12	0.45	6.35	0.01	0.32	0.14-0.78
	NA toward social influence of robots	−0.14	0.39	0.13	0.72	0.87	0.40-1.88
	NA toward emotions in interactions with robots	1.15	0.39	8.94	0.003	3.17	1.49-6.77
Human relationship	Constant	0.39	1.30	0.09	0.77		
	NA toward situations and interactions with robots	−1.38	0.51	7.33	0.007	0.25	0.09-0.68
	NA toward social influence of robots	0.42	0.44	0.91	0.34	1.52	0.64-3.63
	NA toward emotions in interactions with robots	0.74	0.42	3.05	0.08	2.09	0.91-4.79

Note: N=344; Nagelkerke pseudo R^2=0.192.
[a]The reference group is self-determination.
[b]df=1.
OR, odds ratio; NA, negative attitude.

ed toward mechanical improvements at the expense of characteristics that would make the robot more similar to humans.

On the contrary, the self-determination group is mainly fascinated by the implementation of robots as full-fledged social partners who can become members of our society. However, their expectations are unrealistic as they believe that the technical improvements of robots are not particularly relevant to achieving complex social skills. Although this consideration requires further research, it is possible to speculate that the unrealistic expectations of the self-determination group are due to people's limited direct experience with robots.

In the middle of the continuum of humanization of robots, three other profiles have been identified: fuzzy, human attributes, and human relationship. The fuzzy group is generally not interested in development of specific robot features, plausibly representing the people least interested in robotic technologies.

Concerning the groups, human attributes and human relationships, both are interested in human-like performative improvements of robots. However, the group human attributes expects to have robots that will have some human attributes (human life factor higher than the mean, but not as high as in the self-determination group), but not the needed skills (i.e., medium levels of technical features, social-emotional resonance, and agency) to became really like a human. Vice versa, the fourth profile (i.e., human relationship) includes emerging adults who have more realistic expectations; in addition to the desire to enhance human characteristics (the human life factor score is higher than average), the members of this group also expect to improve robots in terms of social-emotional resonance and agency. A combination of human life skills, social-emotional resonance, and agency can result in a robot that is effectively able to interact as a human.

Another important consideration concerns the greater number of people who belong to the group of mechanical properties than other groups; this number suggests that the greatest portion of emerging adults is not expecting to interact with human robots in their future. In general, the greater tendency of people to prefer the mechanical improvement of robots is in line with the Eurobarometer (2012) showing that the majority of the European population considers robots as tools that can support people (e.g., help people in hard or dangerous activities such as manufacturing, search and rescue, security, and so on) and not as peers of human beings.

Unexpectedly, people's expectations of the next generation of robots are independent of the sociodemographic features. This finding could be explained by the high interest that the Italian media took in robotic technologies in recent years. The wide dissemination of information

about robots and their use may have made emerging adults sufficiently aware of these technologies (Sundar et al., 2016) despite their gender, age, education level, and occupational status.

The wide dissemination of information about robots can also explain why the knowledge that people have about robots does not seem to produce any difference in their expectations. The only aspect of knowledge that makes the difference is the definition that people give to the robot. Definitions are related to the representation that people have of the robot (e.g., supportive vs. dangerous tool) and consequently to their attitudes toward robots.

Our study indeed shows how negative attitudes toward robots are strongly related to people's expectations. The self-determination group was identified as a benchmark regarding negative attitudes toward robots because we hypothesized that these individuals, desiring the most humanized robot, should have also expressed fewer negative attitudes compared with the other groups.

However, the data revealed a more complex scenario than initially assumed. The fewer negative attitudes of the groups, mechanical properties, fuzzy, human attributes, and human relationships, with respect to situations and interactions with robots could be explained by their desire to have robots that are efficient and useful, that is, more performative, independently of their purposes of use. In other words, these groups differ from the self-determination group as they are projected toward a future in which robots will be at the service of humans and under their control.

On the contrary, the members of the self-determination group are more worried about situations and interactions with robots compared with the others because their expectations are oriented toward autonomous human-like robots. Therefore, these people are more frightened of situations and interactions because they believe that robots will experience the complexity of human relational dynamics and, consequently, be out of human control. This hypothesis, although speculative, seems to be close to the experience lived by the protagonist of Ian McEwan's recent novel *Machines Like Me* (2019) in which he is not only fascinated by the humanity of the android, named Adam, but also deeply frightened by his unpredictability in relationships.

However, to better clarify this picture of negative attitudes toward robots, it is important to also consider attitudes toward emotions. The results concerning negative emotional attitudes in the groups, mechanical properties, fuzzy, and human attributes, could be plausibly explained by their general concern for the humanization of robots and therefore by their greater relational autonomy, which on the contrary is desired by the groups, human relationship and self-determination.

Moreover, although the self-determination group is not particularly worried about emotional interactions with robots compared with other groups, it is also the most frightened of interactions. This apparently controversial result could be explained by the distance between expectations and desires: on the one hand, people in this group require the highest level of sophistication for robots because they are fascinated by the idea of a robot with human characteristics; on the other hand, they could be terrified, as in anticipation of the Uncanny Valley phenomenon. Moreover, the members of the human relationship group are more inclined to interact with robots than the other groups because, although they are not the only ones with few negative attitudes, they are those who combine the desire for human life aspects with the other skills (technical features, social-emotional resonance, and agency) necessary to have a functional robot.

In conclusion, this study identified emerging adults' expectations regarding development of future robots, detecting five profiles that can be placed along a continuum of humanization of robots. Although there is enough information about robots' current capabilities, people are still concerned about the role that robots will play in society. Therefore, it is important to design interactions between humans and robots, calibrated on the basis of the identified expectation profiles, to support people to overcome their negative attitudes and false myths/false expectations.

Appendix A1 - *Seventeen-item scale for robotic needs*

Questions (Italian)	*Questions (English)*	*Dimension*
Quanto ritieni importante per un generico robot ottenere:	How important do you think it is for a generic robot to:	
1. Maggiore energia (es. durata della batteria)	Have more power (e.g., battery life)	Technical features
2. Sistema di autoregolazione interna (es. sistema di raffreddamento in risposta ad aumento di calore interno)	Have an internal self-regulation system (e.g., cooling system in response to increased internal heat)	
3. Processori, espansioni e potenziamenti hardware o software di vario tipo	Have processors, expansions, and various types of hardware or software upgrades	
4. Un aumento di capacità comunicative	Increase its communication skills	Social-emotional resonance
5. Un aumento di capacità di trasmettere e comunicare emozioni	Increase its ability to transmit and communicate emotions	
6. Un aumento di capacità di riconoscere e provare emozioni	Increase its ability to recognize and experience emotions	

(*segue*)

Questions (Italian)	*Questions (English)*	*Dimension*
7. Capacità di interagire con l'ambiente (es. mani più abili)	Have the ability to interact with the environment (e.g., highly skilled hands)	Agency
8. Essere dotato di mansioni maggiori e più efficaci sensori specifici per l'adattamento a specifici ambienti	Be equipped to perform highly skilled tasks and have very effective sensors for adaptation to specific environments	
9. Poter svolgere compiti che prevedano il coinvolgimento di altre persone o robot	Have the ability to perform tasks with other people or robots	
10. Avere un compagno/a, amico/a	Have a friend/partner	Human life
11. Avere un animale domestico	Have a pet	
12. Denaro	Have money	
13. Poter viaggiare	Have the potential to travel	
14. Avere del tempo libero	Have free time	
15. Poter ascoltare musica	Have the ability to listen to music	
16. Fare sport	Play sports	
17. Fare sesso	Have sex	

Note: Using a 5-point rating scale (1=not at all; 2=little; 3=quite; 4=very; and 5=very much), respondents show how important they think it is for a future robot to achieve the following features.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723.

Arnett, J. J. (2000). Emerging adulthood: A theory of development from the late teens through the twenties. *American psychologist, 55*(5), 469-480.

Bartneck, C., Suzuki, T., Kanda, T., & Nomura, T. (2007). The influence of people's culture and prior experiences with Aibo on their attitude towards robots. *Ai & Society, 21*(1-2), 217-230.

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science robotics, 3*(21).

Bemelmans, R., Gelderblom, G. J., Jonker, P., & De Witte, L. (2012). Socially assistive robots in elderly care: a systematic review into effects and effectiveness. *Journal of the American Medical Directors Association, 13*(2), 114-120.

Billing, E., Rosén, J., & Lindblom, J. (2019). Expectations of robot technology in welfare. In *The second workshop on social robots in therapy and care in conjunction with the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2019)* (pp. 11-14).

Di Dio, C., Manzi, F., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., & Marchetti, A. (2020a). It does not matter who you are: Fairness in pre-schoolers interacting with human and robotic partners. *International Journal of Social Robotics*, *12*(5), 1045-1059.

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., & Marchetti, A. (2020b). Come i bambini pensano alla mente del robot: il ruolo dell'attaccamento e della Teoria della Mente nell'attribuzione di stati mentali ad un agente robotico [How children think about the robot's mind. The role of attachment and Theory of Mind in the attribution of mental states to a robotic agent]. *Sistemi Intelligenti*, *1*(20), 41-56.

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., & Marchetti, A. (2020c). Shall I trust you? From child-robot interaction to trusting relationships. *Frontiers in psychology*, *11*, 469.

Eurobarometer. (2012). *Public attitudes towards robots (Special Eurobarometer 382)*. European Commission. http://ec.europa.eu/public_opinion/archives/ebs/ebs_382_en.pdf.

Kuhnert, B., Ragni, M., & Lindner, F. (2017, August). The gap between human's attitude towards robots in general and human's expectation of an ideal everyday life robot. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 1102-1107). IEEE.

Kwon, M., Jung, M. F., & Knepper, R. A. (2016, March). Human expectations of social robots. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 463-464). IEEE.

Leite, I., Martinho, C., & Paiva, A. (2013). Social robots for long-term interaction: A survey. *International Journal of Social Robotics*, *5*(2), 291-308.

Lin, C. H., Liu, E. Z. F., & Huang, Y. Y. (2012). Exploring parents' perceptions towards educational robots: Gender and socio-economic differences. *British Journal of Educational Technology, 43*(1), E31-E34.

Loffredo, D., & Tavakkoli, A. (2016). What are European Union public attitudes towards robots. *Systemics, Cybernetics and Informatics, 14*(1), 11-19.

Manzi, F., Di Dio, C., Shoji, I., Takayuki, K., Hiroshi, I., Massaro, D., & Marchetti, A. (2020a). Moral evaluation of Human and Robot interactions in Japanese preschoolers. In B. De Carolis, C. Gena, A. Lieto, S. Rossi & A. Sciutti (Eds.), *Workshop on Adapted intEraction with SociAl Robots*: *Vol. 2724)* (pp. 20-27). CEUR Workshop Proceedings.

Manzi, F., Ishikawa, M., Di Dio, C., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., & Marchetti, A. (2020b). The understanding of congruent and incongruent referential gaze in 17-month-old infants: An eye-tracking study comparing human and robot. *Scientific Reports*, *10*(1), 1-10.

Manzi, F., Peretti, G., Di Dio, C., Cangelosi, A., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., & Marchetti, A. (2020c). A robot is not worth another: Exploring chil-

dren's mental state attribution to different humanoid robots. *Frontiers in Psychology, 11*, 2011.

Marchetti, A., Di Dio, C., Manzi, F., & Massaro, D., (2020d). Robotics in clinical and developmental psychology. *Reference Module in Neuroscience and Biobehavioral Psychology*. 2022:B978-0-12-818697-8.00005-4.

Manzi, F., Massaro, D., Di Lernia, D., Maggioni, M. A., Riva, G., & Marchetti, A. (2021). Robots are not all the same: Young adults' expectations, attitudes, and mental attribution to two humanoid social robots. *Cyberpsychology, Behavior, and Social Networking, 24*(5), 307-314.

Marchetti, A., Manzi, F., Itakura, S., & Massaro, D. (2018). Theory of mind and humanoid robots from a lifespan perspective. *Zeitschrift für Psychologie, 226*(2), 98-109.

Masyn, K. (2013). Latent class analysis and finite mixture modeling. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods in Psychology* (pp. 551-611). Oxford University Press.

McEwan, I. (2019). *Machines Like Me*. Jonathan Cape.

Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological methods, 4*(2), 139.

Nomura, T., Kanda, T., Suzuki, T., & Kato, K. (2008). Prediction of human behavior in human-robot interaction using psychological scales for anxiety and negative attitudes toward robots. *IEEE Transactions on Robotics, 24*(2), 442-451.

Nomura, T., Suzuki, T., Kanda, T., & Kato, K. (2006). Measurement of negative attitudes toward robots. *Interaction Studies, 7*(3), 437-454.

Riva, G., Wiederhold, B. K., Di Lernia, D., Chirico, A., Riva, E. F. M., Mantovani, F., Cipresso, P., & Gaggioli, A. (2019). Virtual reality meets artificial intelligence: The emergence of advanced digital therapeutics and digital biomarkers. *Annual Review of CyberTherapy and Telemedicine, 17*, 3-7.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461-464.

Sciutti, A., Mara, M., Tagliasco, V., & Sandini, G. (2018). Humanizing human-robot interaction: On the importance of mutual understanding. *IEEE Technology and Society Magazine, 37*(1), 22-29.

Serino, S., Scarpina, F., Dakanalis, A., Keizer, A., Pedroli, E., Castelnuovo, G., Chirico, A., Catallo, V., Di Lernia, D., & Riva, G. (2018). The role of age on multisensory bodily experience: An experimental study with a virtual reality full-body illusion. *Cyberpsychology, Behavior, and Social Networking, 21*(5), 304-310.

Sorgente, A., Lanz, M., Serido, J., Tagliabue, S., & Shim, S. (2019). Latent transition analysis: Guidelines and an application to emerging adults' social development. *TPM: Testing, Psychometrics, Methodology in Applied Psychology, 26*(1).

Sundar, S. S., Waddell, T. F., & Jung, E. H. (2016, March). The Hollywood robot syndrome media effects on older adults' attitudes toward robots and adoption intentions. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 343-350). IEEE.

Sung, J., Grinter, R. E., & Christensen, H. I. (2010). Domestic robot ecology. *International Journal of Social Robotics, 2*(4), 417-429.

Tapus, A., Mataric, M. J., & Scassellati, B. (2007). Socially assistive robotics [grand challenges of robotics]. *IEEE robotics & automation magazine, 14*(1), 35-42.

Thompson, A. G., & Sunol, R. (1995). Expectations as determinants of patient satisfaction: Concepts, theory and evidence. *International journal for quality in health care, 7*(2), 127-141.

Vannucci, F., Sciutti, A., Lehman, H., Sandini, G., Nagai, Y., & Rea, F. (2019). Cultural differences in speed adaptation in human-robot interaction tasks. *Paladyn, Journal of Behavioral Robotics, 10*(1), 256-266.

3. Effect of Social Anxiety on the Adoption of Robotic Training Partner

D.H. Zhu, Z.Z. Deng

ABSTRACT

With the development of artificial intelligence technologies, robotic training partner is becoming a reality, which is a substitute for human training partner. Socially anxious individuals feel uncomfortable in front of unfamiliar people or when being observed by others. Playing with robotic training partners can avoid face-to-face interaction with other people. It is unclear whether social anxiety affects the adoption of robotic training partners. This study investigates the effect of social anxiety on the adoption of robotic training partners among university students. Study 1 confirmed that university students with higher social anxiety are more likely to choose robotic training partners than human training partners. Mediation analysis in Study 2 supported the mediating role of sense of relaxation with robotic training partner in the positive effect of social anxiety on the adoption of robotic training partner. This study shows that developing training partner robots is a meaningful thing for corporate profits and the health of socially anxious people.

Introduction

With the development of artificial intelligence technologies, more and more industries, such as hotels, restaurants, and sports, are developing intelligent physical robots to replace human beings (Karunarathne et al., 2019; Zhu & Chang, 2020). In the field of sports, the table tennis robot named Forpheus developed by Omron is able to practice table tennis with amateurs and help professional athletes to improve their levels by playing on a real table tennis table. It can remember the characteristics of players and judge the player's table tennis level by the accu-

---

This chapter was originally published as Zhu, D.H., & Deng, Z.Z. (2021). Effect of social anxiety on the adoption of robotic training partner. *Cyberpsychology, Behavior, and Social Networking, 24*(5), 343-348. This work was supported by the National Natural Science Foundation of China (Grant Nos. 71972080 and 71720107004).

mulated body information, the way of swinging the racket and the ball track information of the player, and then changing its way of returning the ball to choose a more suitable way to match the level of the player. In the badminton field, there is already a badminton robot named Kengoro developed in Japan and a badminton robot named Robomintoner developed in China. These sports require at least two people to play together before the advent of a robotic training partner. The advent of robotic training partners enables people to do the same physical exercises in the absence of other people.

Socially anxious individuals are shy and feel uncomfortable when they are in the presence of unfamiliar people or observed by other people (Velting & Albano, 2001). They often avoid face-to-face interactions to escape their social fears (Peter et al., 2005). Taking playing table tennis as an example, when playing it with a human training partner, socially anxious individuals may feel their table tennis skills checked by unfamiliar people, which may reduce their interest in playing table tennis. Robotic training partners allow them to play table tennis without other people, which may be a useful tool to reduce social anxiety and increase exercise. However, previous studies show that many people have low willingness to accept robots for the subjective perception of robots lacking emotion (Konok et al., 2018; Fuchs et al., 2015) and the feeling of discomfort with humanoid robots (Ho & McDorman, 2010; Mende et al., 2019). Knowledge about the effect of social anxiety on the adoption of robotic training partners is scarce.

This study aims to examine whether social anxiety will affect the adoption of robotic training partners. Taking part in physical exercise is an effective way to keep healthy and relax. More and more enterprises are investing in the development of intelligent sports robots. If playing with robotic training partners could promote the physical exercise of people with social anxiety, it is meaningful to develop training partner robots for the benefit of enterprises and the health of people with social anxiety.

Literature Review and Hypotheses

Social anxiety and the adoption of robotic training partner

Social anxiety has received considerable attention in social psychology and personality research. Social anxiety is “a state of anxiety resulting from the prospect or presence of interpersonal evaluation in real or imagined social settings” (Leary, 1983, p. 67). Anderson and Harvey suggest that social anxiety may originate from unpleasant social experi-

ences with peers during childhood and adolescence (Anderson & Harvey, 1988). Studies have shown that socially anxious individuals focus on the 'observer's perspective' on themselves and want to create a positive self-image in others along with a lack of confidence in self-presentation, which triggers a fear of negative evaluation and then forms anxiety (Schlenker & Leary, 1985; Spurr & Stopa, 2003).

Because anxiety is an uncomfortable experience, socially anxious people have motivation to minimize their anxiety. As nonverbal cues in face-to-face interactions that usually attract their attention frequently trigger social anxiety (Ko et al., 2014; Veit et al., 2002), those with social anxiety are afraid of face-to-face interactions, which leads to them often avoiding face-to-face interactions to escape social fears (Peter et al., 2005; Pierce, 2009). As such, novel technologies that can help avoid face-to-face interactions have been welcomed by people with social anxiety. For example, Papacharissi and Rubin (2000) find that those with social anxiety tend to use the Internet more as the medium of social interactions, while those who like face-to-face interactions tend to use the Internet more as the tool of information searches. Valkenburg and Peter (2009) show that online communications decrease social anxiety by avoiding the nonverbal cues of face-to-face interactions. Becker and Pizzutti (2017) find that consumers with higher social anxiety are more satisfied with C2C interaction in online shopping than offline shopping. Lin and colleagues (2017) show that social network sites positively affect the recovery of people with high social anxiety after social exclusion, but hinder the recovery of people with low social anxiety.

Robotic training partner is another novel technology to help players avoid face-to-face interactions. The difference between the robotic training partner and the abovementioned Internet, online communications, online shopping, and social network technologies is that the latter is only a medium between socially anxious people and other people, while the former is a substitute for other people. As a substitute, it is not clear whether people with social anxiety are willing to use it. As the need to decrease anxiety motivates those with social anxiety to minimize the chances of being observed by others (Caplan, 2006) and playing with robotic training partners can avoid being observed by human training partners, this study proposes that socially anxious people are more likely to use sparring robots. The hypothesis is as follows:

H1: People with higher social anxiety are more likely to adopt robotic (vs. human) training partner.

Sense of relaxation with robotic training partner as a mediator

Socially anxious individuals want to build a positive self-image in others and fear negative evaluation (Schlenker & Leary, 1985; Spurr & Stopa, 2003). They feel uncomfortable when they are in the presence of unfamiliar people or observed by others (Velting & Albano, 2001). Taking playing table tennis as an example, those with social anxiety may be afraid of negative evaluation from human training partners on their skills and performance, which may decrease their interest in playing table tennis. Making socially anxious individuals feel relaxed is an effective way to reduce social anxiety (Gould et al., 2019; Lu et al., 2020). When the training partner is a robot rather than an unfamiliar person, as a robot is a machine lacking in feeling, even if they have poor performance while playing table tennis, they may feel more relaxed playing with a robotic training partner than with a human training partner. Their anxiety of a negative evaluation from the partner will be mitigated. As the sense of relaxation can help reduce the social anxiety that prevents them from playing table tennis, they may be more likely to choose a robotic training partner. Therefore, this study proposes that the sense of relaxation with robotic training partner makes socially anxious people more willing to adopt a robotic training partner than a human training partner. The hypothesis is as follows:

H2: Social anxiety positively affects the adoption of robotic training partner through the mediator of sense of relaxation with robotic training partner.

Study 1

The purpose of Study 1 is to test whether people with higher social anxiety are more likely to choose a robotic training partner in the presence of human and robotic training partner.

Research design

Given that university students more easily accept new technology and social anxiety is common among them (Turner et al., 1991), the participants were 100 university students recruited by using a professional online survey site in China with monetary reward. They filled in an online questionnaire on the site. The participants were aged between 18 and 25 years (M=22) and 69% of them were females.

The participants first read a paragraph of background material about the development of artificial intelligence technologies and the advent of

various table tennis robots with some dynamic pictures of human/robot playing table tennis. Next, they were asked to read the second material: 'Please imagine that: you have signed up for a table tennis program at a training center to keep healthy. The program can provide each player with a separate indoor room and a sparring partner. This is a professional training center that offers human and robotic sparring partners. The robotic sparring partner introduced by the center has been in operation for 6 months. Professional assessment and customer reviews indicate that the robotic sparring partner is as professional as human sparring. Professionalization means that training partner can change the way of returning the ball to choose a more suitable way to match the level of the player, so that players can enjoy playing. Before starting the project, you need to choose the type of sparring partner'.

After reading the material, attention tests were taken. Next, the participants were asked to fill in a questionnaire about the measurement items of choice intention, social anxiety, table tennis liking, gender, and age. Six participants who failed the attention test were excluded, and the effective sample was 94.

Measurement

Choice intention. Choice intention was measured with a scale adapted from Phau et al. (2010). The scale included three items, such as 'I prefer to choose human training partner/robotic training partner' and 'I want to choose human training partner/robotic training partner'. The participants were asked to choose between human training partner and robotic training partner on a 7-point scale. The option of human training partner was on the far left and the option of robotic training partner was on the far right of the scale.

Social anxiety. Social anxiety was measured by the scale developed by Apaolaza et al. (2019). It includes six items that focus on interpersonal concern in actual interactions from the Social Anxiousness Scale (Anderson & Harvey, 1988; Leary & Kowalski, 1993), such as 'I often feel nervous even in causal get-togethers' and 'I often feel nervous when calling someone on the phone I don't know very well'. The items of social anxiety were tested by using a 7-point Likert scale with 1 = very rarely and 7 = very often.

Descriptive statistics and the Cronbach's alpha value of the two variables are shown in Table 1. The Pearson correlation coefficients among choice intention, social anxiety, table tennis liking, and gender are shown in Table 2.

Table 1 - *Variables and measurement items of Study 1 and Study 2*

	Study 1			*Study 2*		
Item	*Mean*	*SD*	*α*	*Mean*	*SD*	*α*
Social anxiety	4.824	1.189	0.894	4.946	1.133	0.875
I often feel nervous even in casual get-togethers						
I often feel nervous when calling someone on the phone I don't know very well						
I sometimes feel tense when talking to people if I don't know them very well						
I often feel nervous when talking to a person I feel attracted to						
I get nervous when I speak to someone in a position of authority						
In general, I am a shy person						
Choice intention	4.316	1.585	0.937			
I prefer to choose (human training partner——robotic training partner)						
I intent to choose (human training partner——robotic training partner)						
I want to choose (human training partner——robotic training partner)						
Adoption intention				4.698	1.413	0.954
The likelihood of adopting robotic training partner is (very low——very high)						
The probability that I would consider adopting robotic training partner is (very low——very high)						
My willingness to adopt robotic training partner is (very low——very high)						
Sense of relaxation with robotic training partner				5.181	1.457	0.950
Playing with robotic training partner helps me allay fears about my table tennis skills						
Playing with robotic training partner helps me allay fears about making too many mistakes on serve and return						
Robotic training partner makes me feel at ease with playing table tennis						

Table 2 - *Pearson Correlation Coefficients in Study 1 and Study 2*

Variable	*Study 1*				*Study 2*				
	1	*2*	*3*	*4*	*1*	*2*	*3*	*4*	*5*
1. Gender	1.000				1.000				
2. Table tennis liking	–0.014	1.000			–0.183*	1.000			
3. Social anxiety	0.062	–0.092	1.000		–0.010	–0.079	1.000		
4. Adoption intention[a]	–0.133	–0.154	0.339***	1.000	–0.003	–0.062	0.237**	1.000	
5. Sense of relaxation					–0.013	–0.071	0.461***	0.657***	1.000

*$p<0.05$, **$p<0.01$, ***$p<0.001$.
[a] 'Choice intention' in Study 1.

Result

As this study predicted, participants with higher social anxiety would be more likely to adopt robotic (vs. human) training partner, a linear regression analysis with social anxiety as the independent variable, choice intention as the dependent variable, and table tennis liking and gender as the control variables were used to test the influence of social anxiety on choice intention toward a robotic training partner. The result showed that social anxiety had a significantly positive effect on the choice intention of robotic training partner ($\beta=0.337$, $p=0.001$), while the control variables of table tennis liking ($\beta=-0.125$, $p>0.1$) and gender ($\beta=-0.156$, $p>0.1$) had no significant effect (Table 3). By averaging

Table 3 - *Regression analysis for choice/adoption intention toward robotic training partner*

Variable	*Study 1*		*Study 2*	
	β	*t*	*B*	*t*
Control variable				
Gender	–0.156[n]	–1.603	–0.045[n]	–0.095
The degree of liking for table tennis	–0.125[n]	–1.283	–0.008[n]	–0.505
Independent variable				
Social anxiety	0.337***	3.452	0.233**	2.652

**$p<0.01$, ***$p<0.001$.
n, not significant.

the items of social anxiety and taking the mean value higher than four as the high social anxiety group and the others as the low social anxiety group, a one-way analysis of variance (ANOVA) confirmed that the high social anxiety group had higher intention to choose a robotic training partner than the low social anxiety group [$M_{high}=4.60$, $SD=1.54$; $M_{low}=3.33$, $SD=1.41$; $F(1, 92)=11.407$, $p=0.001$].

The results of Study 1 show that university students with higher social anxiety are more likely to adopt a robotic training partner. Hypothesis 1 was supported.

Study 2

The purpose of Study 2 was to test whether people with higher social anxiety are more likely to adopt robotic training partners because they think playing with robotic training partners makes them have a sense of relaxation when exercising.

Research design

There were 150 university students recruited by using a professional online survey site in China with monetary reward. The participants were aged between 18 and 25 years ($M=22$) and 72% were females. First, the participants' social anxiety was tested. Next, the participants were asked to read the same background material about the introduction of artificial intelligence technologies and table tennis robots and the table tennis program material used in Study 1. After attention tests, the participants were asked to fill in the measurement items of robotic training partner adoption intention, sense of relaxation with a robotic training partner, table tennis liking, gender, and age. Twenty-three participants who failed the attention tests were excluded, with a valid sample of 127.

Measurement

Social anxiety. Social anxiety was measured with the six items used in Study 1.

Adoption intention. Adoption intention was measured with a scale adapted from Zhu et al. (2014). The scale included three items, such as 'The likelihood of adopting a robotic training partner is (very low/very high)' and 'My willingness to adopt a robotic training partner is (very

low/very high).' The items were tested by using a 7-point Likert scale with 1 = very low and 7 = very high.

Sense of relaxation with robotic training partner. Sense of relaxation with a robotic training partner was measured with a scale developed by the current study. The scale included three items, such as 'Playing with a robotic training partner helps me allay fears about my table tennis skills' and 'Robotic training partner makes me feel at ease with playing table tennis'. The items were tested by using a 7-point Likert scale with 1 = strongly disagree and 7 = strongly agree.

Descriptive statistics and the Cronbach's alpha value of the three variables are shown in Table 1. The Pearson correlation coefficients among adoption intention, sense of relaxation with robotic trainer partner, social anxiety, table tennis liking, and gender are shown in Table 2.

Result

Adoption intention. As this study predicted, social anxiety positively affects the adoption of robotic training partner, a linear regression analysis with social anxiety as the independent variable, adoption intention as the dependent variable, and table tennis liking and gender as the control variables were used to test the influence of social anxiety on the adoption intention of robotic training partner. The result showed that social anxiety had a significantly positive effect on adoption intention toward a robotic training partner ($\beta = 0.233$, $p = 0.009$), while the control variables of table tennis liking ($\beta = -0.008$, $p > 0.1$) and gender ($\beta = -0.045$, $p > 0.1$) had no significant effect (Table 3). A one-way ANOVA confirmed that the high social anxiety group had the higher adoption intention of robotic training partner than the low social anxiety group [$M_{high} = 5.38$, $SD = 0.73$; $M_{low} = 3.24$, $SD = 0.74$; $F(1, 125) = 176.526$, $p < 0.001$]. Hypothesis 1 was supported again.

Sense of relaxation with robotic training partner as the mediator. A one-way ANOVA showed that the high social anxiety group had a higher sense of relaxation with a robotic training partner than the low social anxiety group [$M_{high} = 5.37$, $SD = 1.30$; $M_{low} = 4.45$, $SD = 1.80$; $F(1, 125) = 8.769$, $p = 0.004$]. To examine whether the sense of relaxation with a robotic training partner would mediate the relationship between social anxiety and adoption intention, the mediation analysis procedure (model 4) recommended by Hayes (2017) was used. The mediation effect was tested using 5,000 bootstrap samples, with social anxiety as the independent variable, sense of relaxation with robotic train-

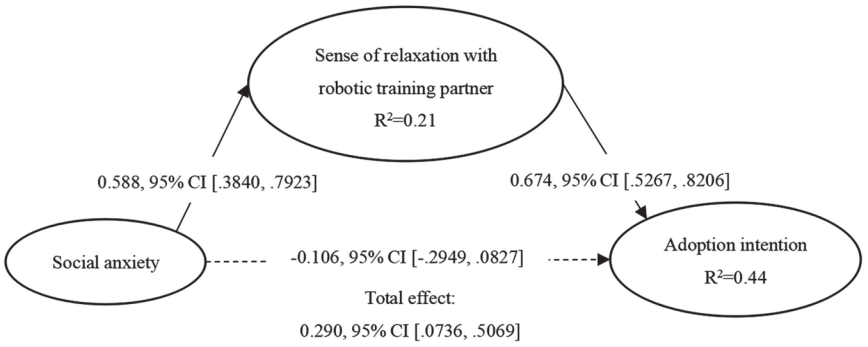
ing partner as the mediator, adoption intention as the dependent variable, and table tennis liking and gender as covariates. As predicted, the 95% confidence interval (CI) did not include zero (0.396, 95% CI [0.2136, 0.5825]), which demonstrated that sense of relaxation with robotic training partner mediates the relationship between social anxiety and adoption intention (Figure 1).

Discussion and Implications

This study verified that socially anxious individuals are more likely to adopt a robotic training partner and the sense of relaxation with robotic training partner is the mediator. Social anxiety stems from people's fear of face-to-face interactions. Robotic training partner is a substitute for human training partner and it is essentially a machine lacking in feeling. Hence, socially anxious people can feel relaxed when doing physical exercise with a robotic training partner. As such, socially anxious individuals are more likely to adopt a robotic training partner. The findings suggest that a robotic training partner is a useful new technology that can help socially anxious people reduce social anxiety and take part in physical exercise.

This is the first study to investigate the effect of social anxiety on the adoption of robots. Previous studies show that many people have a low willingness to accept robots (Konok et al., 2018; Fuchs et al., 2015; Ho & MacDorman, 2010; Mende et al., 2019). On the contrary, this study finds that people with social anxiety have a high willingness to adopt robots and reveals that the effect is mediated by the sense of relaxation with robots, which expands the literature about social anxiety and technology

Figure 1 - *The mediating effect test*



use. Another contribution of this study is to pay attention to the training partner robots. Previous studies mainly focused on social robots and health care robots (Ivanov & Webster, 2019; Zhu & Chang, 2020). Although it has great use value, there is little research on training partner robots in scientific literature. This study expands the types of robots in the literature on robotics.

In practice, the findings of this study suggest that the commercialization of sports robots aimed at servicing amateurs can take those with social anxiety as an important part of key customers. Meanwhile, robotic training partner programs should be aware that it is necessary to offer a separate room for those with social anxiety to avoid being observed by other people. In addition, besides sports robots, robot companies can also design and develop other types of robots by targeting socially anxious people.

This study also has implications for public policy makers. Obesity is an important public health problem in many countries (Hock & Bagchi, 2018). Obese people are more likely to suffer from social anxiety (Abdollahi & Talib, 2015). Taking part in physical exercise is an effective method to keep fit and relaxed. Therefore, public policy makers can build public health gymnasiums and separate rooms equipped with training partner robots to help obese people relieve their psychological pressure of doing exercises with others.

This study has some limitations. First, this study focused on student samples. Although students and nonstudents with social anxiety share a common characteristic of avoiding face-to-face interaction, future studies will need to replicate this study in nonstudent samples to enhance the robustness of the conclusions. Second, limited by the commercial development of robotic training partner at the present stage, this study adopted the scenario-based investigation method. Future research should investigate the real users of robotic training partners. Third, moderating variables should be examined in future studies, such as high-technology preference and the anthropomorphic characteristics of robotic training partner.

References

Abdollahi, A., & Talib, M. A. (2015). Sedentary behaviour and social anxiety in obese individuals: The mediating role of body esteem. *Psychology, health & medicine*, *20*(2), 205-209.

Anderson, C. A., & Harvey, R. J. (1988). Discriminating between problems in living: An examination of measures of depression, loneliness, shyness, and social anxiety. *Journal of social and clinical psychology*, *6*(3-4), 482-491.

Apaolaza, V., Hartmann, P., D'Souza, C., & Gilsanz, A. (2019). Mindfulness, compulsive mobile social media use, and derived stress: The mediating roles of self-esteem and social anxiety. *Cyberpsychology, Behavior, and Social Networking*, *22*(6), 388-396.

Becker, L. C., & Pizzutti, C. (2017). C2C value creation: Social anxiety and retail environment. *Journal of Research in Interactive Marketing*, *11*, 398-415.

Caplan, S. E. (2006). Relations among loneliness, social anxiety, and problematic Internet use. *CyberPsychology & behavior*, *10*(2), 234-242.

Fuchs, C., Schreier, M., & Van Osselaer, S. M. (2015). The handmade effect: What's love got to do with it?. *Journal of marketing*, *79*(2), 98-110.

Gould, C. E., Kok, B. C., Ma, V. K., Wetherell, J. L., Sudheimer, K., & Beaudreau, S. A. (2019). Video-delivered relaxation intervention reduces late-life anxiety: A pilot randomized controlled trial. *The American Journal of Geriatric Psychiatry*, *27*(5), 514-525.

Hayes, A. F. (2017). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach (2nd Ed.)*. Guilford publications.

Ho, C. C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, *26*(6), 1508-1518.

Hock, S. J., & Bagchi, R. (2018). The impact of crowding on calorie consumption. *Journal of Consumer Research*, *44*(5), 1123-1140.

Ivanov, S., & Webster, C. (2019). *Robots, Artificial Intelligence and Service Automation in Travel, Tourism and Hospitality*. Emerald Publishing.

Karunarathne, D., Morales, Y., Nomura, T., Kanda, T., & Ishiguro, H. (2019). Will older adults accept a humanoid robot as a walking partner?. *International Journal of Social Robotics*, *11*(2), 343-358.

Ko, C. H., Liu, T. L., Wang, P. W., Chen, C. S., Yen, C. F., & Yen, J. Y. (2014). The exacerbation of depression, hostility, and social anxiety in the course of Internet addiction among adolescents: A prospective study. *Comprehensive Psychiatry*, *55*(6), 1377-1384.

Konok, V., Korcsok, B., Miklósi, Á., & Gácsi, M. (2018). Should we love robots? – The most liked qualities of companion dogs and how they can be implemented in social robots. *Computers in Human Behavior*, *80*, 132-142.

Leary, M. R. (1983). Social anxiousness: The construct and its measurement. *Journal of personality assessment*, *47*(1), 66-75.

Leary, M. R., & Kowalski, R. M. (1993). The interaction anxiousness scale: Construct and criterion-related validity. *Journal of personality assessment*, *61*(1), 136-146.

Lin, X., Li, S., & Qu, C. (2017). Social network sites influence recovery from social exclusion: Individual differences in social anxiety. *Computers in Human Behavior*, *75*, 538-546.

Lu, S. M., Lin, M. F., & Chang, H. J. (2020). Progressive muscle relaxation for patients with chronic schizophrenia: A randomized controlled study. *Perspectives in psychiatric care*, *56*(1), 86-94.

Mende, M., Scott, M. L., van Doorn, J., Grewal, D., & Shanks, I. (2019). Service robots rising: How humanoid robots influence service experiences and elicit compensatory consumer responses. *Journal of Marketing Research*, *56*(4), 535-556.

Papacharissi, Z., & Rubin, A. M. (2000). Predictors of Internet use. *Journal of broadcasting & electronic media*, *44*(2), 175-196.

Peter, J., Valkenburg, P. M., & Schouten, A. P. (2005). Developing a model of adolescent friendship formation on the Internet. *CyberPsychology & Behavior*, *8*(5), 423-430.

Phau, I., Shanka, T., & Dhayan, N. (2010). Destination image and choice intention of university student travellers to Mauritius. *International Journal of Contemporary Hospitality Management*, *22*, 758-764.

Pierce, T. (2009). Social anxiety and technology: Face-to-face communication versus technological communication among teens. *Computers in Human Behavior*, *25*(6), 1367-1372.

Schlenker, B. R., & Leary, M. R. (1985). Social anxiety and communication about the self. *Journal of Language and Social Psychology*, *4*(3-4), 171-192.

Spurr, J. M., & Stopa, L. (2003). The observer perspective: Effects on social anxiety and performance. *Behaviour Research and Therapy*, *41*(9), 1009-1028.

Turner, S. M., Beidel, D. C., Borden, J. W., Stanley, M. A., & Jacob, R. G. (1991). Social phobia: Axis I and II correlates. *Journal of Abnormal Psychology*, *100*(1), 102.

Valkenburg, P. M., & Peter, J. (2009). Social consequences of the Internet for adolescents: A decade of research. *Current directions in psychological science*, *18*(1), 1-5.

Veit, R., Flor, H., Erb, M., Hermann, C., Lotze, M., Grodd, W., & Birbaumer, N. (2002). Brain circuits involved in emotional learning in antisocial behavior and social phobia in humans. *Neuroscience letters*, *328*(3), 233-236.

Velting, O. N., & Albano, A. M. (2001). Current trends in the understanding and treatment of social phobia in youth. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(1), 127-140.

Zhu, D. H., & Chang, Y. P. (2020). Robot with humanoid hands cooks food better? Effect of robotic chef anthropomorphism on food quality prediction. *International Journal of Contemporary Hospitality Management*, *32*, 1367-1383.

Zhu, D. H., Chang, Y. P., Luo, J. J., & Li, X. (2014). Understanding the adoption of location-based recommendation agents among active users of social networking sites. *Information Processing & Management*, *50*(5), 675-682.

4. The Understanding of Congruent and Incongruent Referential Gaze in 17-Month-Old Infants

An Eye-Tracking Study Comparing Human and Robot

F. Manzi, M. Ishikawa, C. Di Dio, S. Itakura, T. Kanda, H. Ishiguro, D. Massaro, A. Marchetti

ABSTRACT

Several studies have shown that the human gaze, but not the robot gaze, has significant effects on infant social cognition and facilitate social engagement. The present study investigates early understanding of the referential nature of gaze by comparing – through the eye-tracking technique – infants' response to human and robot's gaze. Data were acquired on thirty-two 17-month-old infants, watching four video clips, where either a human or a humanoid robot performed an action on a target. The agent's gaze was either turned to the target (congruent) or opposite to it (incongruent). The results generally showed that, independent of the agent, the infants attended longer at the face area compared to the hand and target. Additionally, the effect of referential gaze on infants' attention to the target was greater when infants watched the human compared to the robot's action. These results suggest the presence, in infants, of two distinct levels of gaze-following mechanisms: one recognizing the other as a potential interactive partner, the second recognizing partner's agency. In this study, infants recognized the robot as a potential interactive partner, whereas ascribed agency more readily to the human, thus suggesting that the process of generalizability of gazing behaviour to non-humans is not immediate.

This chapter was originally published as Manzi, F., Ishikawa, M., Di Dio, C., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., & Marchetti, A. (2020). The understanding of congruent and incongruent referential gaze in 17-month-old infants: An eye-tracking study comparing human and robot. *Scientific Reports, 10*, 11918. This research was funded by the Japanese Society for the Promotion of Science, Programme Grant # 16H01880, 16H06301, 15H01846, 25245067. Also, this research was supported by JST CREST Grant-Number JPMJCR17A2, Japan. Finally, the research received a support by the Università Cattolica del Sacro Cuore, Milan. In addition, we would like to thank the entire Baby Science Centre of Doshisha University and Kyoto University, especially Fumina Sano and Nanami Toya, for data collection. F.M. and M.I. conceived the experiment. M.I. conducted the experiment. F.M., M.I. and C.D.D. analysed the results. All authors contributed to the writing of the manuscript. The author(s) declare no competing interests.

Introduction

Human infants display a sensitivity towards the human eyes from birth, a phenomenon known as preference for direct gaze (Farroni et al., 2002). Direct gaze has been suggested to activate social brain networks and facilitate social cognition in infants (Senju & Johnson, 2009). Additionally, some studies have shown that averted gaze also affects infants' cognitive processing of objects and faces (Ishikawa & Itakura, 2018; Ishikawa et al., 2019). It has been shown that new-born babies can discriminate between direct and averted gaze as they are faster to make saccades to peripheral targets cued by direct gaze (Farroni et al., 2004). Such sensitivity to other's eye-gaze may be regarded as precursor of gaze following in later development, which is essential for efficient social learning (Csibra & Gergely, 2009). In experimental settings, gaze following has been studied in infants as young as 3 months (D'Entremont et al., 1997). Data generally show that infants begin to follow eye-gaze at about 6 months (Butterworth & Jarrett, 1991; Senju & Csibra, 2008; Gredebäck et al., 2010). Infants have also shown to be able to use information provided by eye-gaze to understand intentions behind actions. For example, it has been shown that infants use other's gaze direction to predict or anticipate action (Phillips et al., 2002). In the study by Phillips et al. (2002), an actor grasped one of two objects in a situation where cues from the actor's gaze could serve to determine which object would be grasped. The results showed that 12- and 14-month-olds, but not 8-month-olds, recognized that the actor was likely to grasp the object which she looked at before grasping, in line with the referential nature of the actor's gaze.

In the present study, we explored early understanding of referential gaze toward a target when infants observed either a human or a humanoid robot performing an action toward an object in a gaze-following task (Slaughter & McConnell, 2003). This task resembles the classical visuospatial cueing paradigm (Posner, 1980) used to evaluate the process of attention shift, i.e. visuo-spatial orienting (Driver et al., 1999; Friesen & Kingstone, 1998; Langdon & Smith, 2005), toward the same direction/object/event that another person is attending to. Using a gaze-following task, Senju et al. (2006) studied the referential gaze with infants. The authors recorded ERPs of adults and 9-month-old infants while watching scenes containing a human congruent-object gaze shifts or incongruent-object gaze shifts. The ERP results suggest that 9-month-olds process the object-congruent gaze shifts similarly to adults. In addition, an early frontal component was observed in infants, which showed greater amplitude in response to congruent gaze. This component may reflect a rapid processing of socially relevant information, such as the identification of communicative or informative situations, and could provide a

basis for the development of attention sharing, social learning and Theory of Mind (ToM).

With respect to robots, several studies have shown that people tend to perceive robots as interactive partners (Marchetti et al., 2018), although – with increasing age – generally attribute to them poor human-like mental qualities (Di Dio et al., 2019, 2020a,b). Only a few studies, however, have investigated early socio-cognitive responses to robots' behaviour (Okumura et al., 2013a,b). While the human gaze has been shown to have significant effects on infants' social cognition and facilitate social engagement (Ishikawa et al., 2019; Ishikawa & Itakura, 2019), social cognitive studies with robots have failed to find an effect of the robot gaze on infants' behaviour. For example, while reporting that 12-months-old infants do follow the robot gaze direction, Okumura et al. (2013a) also found that the robot gaze does not facilitate the learning processes of a new object. Nevertheless, infants appear to become sensible to the direction of the robot gaze if they first observe the robot engaging in social interaction with an adult (Meltzoff et al., 2010). Likewise, infants imitate the robot's goal-directed actions only when the robot first establishes eye contact with an adult (Itakura et al., 2008). These findings suggest that the robot gaze may not be as meaningful to the infant as the human gaze, although it can be charged of social meaning through a triadic social engagement, where the human adult provides a role model of the relationship.

Additionally, it has also been shown that infants do not display anticipatory gaze in response to robots' action, suggesting a lack of intentionality attribution to the robot (Okumura et al., 2013b). In support of this observation, Kanakogi and Itakura (Kanakogi & Itakura, 2011) found that 10-month-old infants showed anticipatory looking at human reaching actions, but no anticipation in response to a mechanical hand. Furthermore, a brain imaging study using functional near-infrared spectroscopy (fNIRS) examined infants' brain activation while watching either functional or non-functional actions performed by a human and a mechanical hand (Biondi et al., 2016). The result showed that the left middle-posterior temporal cortex responded selectively to the human hand, but only in the context of functionally relevant actions on objects. This evidences infants' sensitivity to agency when a meaningful action is performed by a human and not by robots.

Altogether, these findings seem to suggest poor involvement of early social cognition when facing a robotic agent. Also, there is no direct evidence of infants' sensitivity to the robot's action when paired with the robot referential gaze, which represents an extremely important cue enhancing social engagement. To this purpose, we monitored the gazing behaviour of 17-month-olds infants while watching video-clips showing

either a human or a humanoid robot performing an action. We selectively involved 17-month-old infants because at this age infants are able to follow the gaze of others (Gredebäck et al., 2008) and have reached an understanding of the referential nature of the gaze (Butterworth & Cochran, 1980; Butterworth & Itakura, 2000; D'Entremont et al., 1997; Hood et al., 1998; Morales et al., 1998; Woodward, 2003). The action performed by the agents was either congruently or incongruently anticipated by the direction of the agent's gaze. Additionally, to enhance socio-cognitive engagement, we also had the human and robot agents either performing or not direct eye-contact to the infant before action performance. Based on previous findings, we hypothesized to find the following: 1) infants would be equally sensitive to the effect of direct eye-contact on attention by both agents; 2) infants would be more sensitive to the referential gaze of both agents in the congruent with respect to the incongruent condition; 3) in the congruent condition, infants would be more sensitive to the referential gaze of the human with respect to the robot agent.

Methods

Participants

Data were obtained from an initial sample of 36 infants. Four infants were excluded from the analyses because of inattentiveness (one, whose gaze pattern was fuzzy; two who completed fewer than 3 trials of gazing; one for technical acquisition errors). The final sample was composed of thirty-two 17-month-old Japanese infants (F=15; M=17.43 months, SD=0.96). The infants were divided into two groups for each condition as follows: 1) 16 infants for the Human condition (F=8; M=17.25 months, SD=.83); 2) 16 infants for the Robot condition (F=7; M=17.64 months, SD=1.08). The children's parents received a written explanation of the procedure of the study, the measurement items, and provided written informed consent before their infants took part in the study. The experimental protocol was approved by the Research Ethics Review Board, Department of Psychology, Kyoto University, Japan.

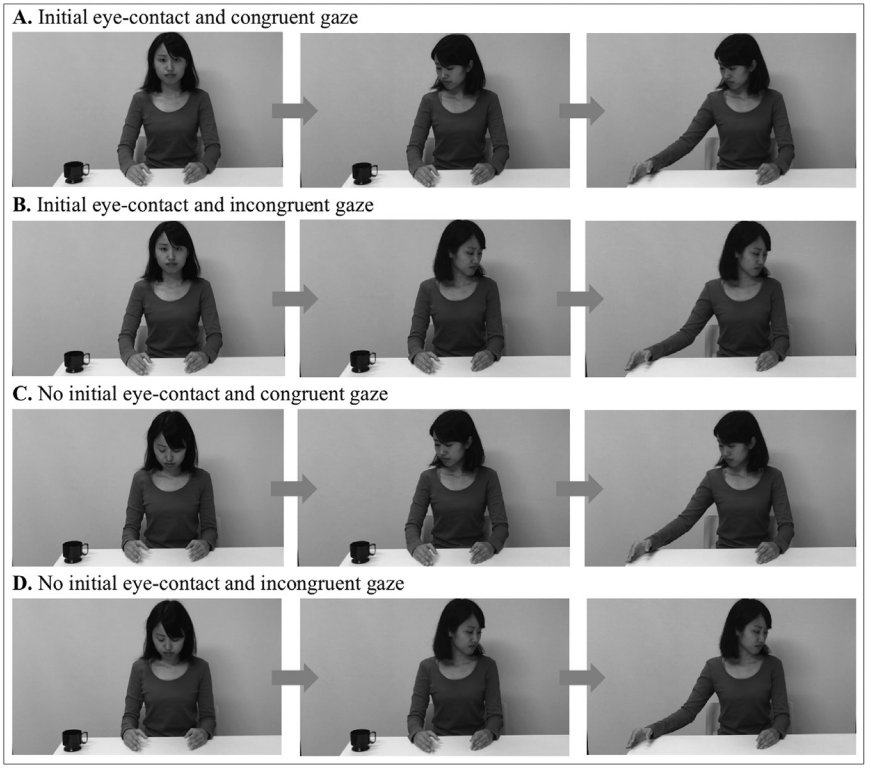
Design and stimuli

The design of the study was a multifactorial 2x2x3x3x2 mixed-model, with 2 levels of initial eye-contact (present, absent), 2 levels of gaze-congruence with the action (congruent, incongruent), 3 levels of ar-

ea of interests (AOIs: face, hand, object), 3 levels of video sequence (no movement, head movement, hand movement) as the within-subject factors, and 2 levels of agency (Human, Robot) as the between-subject factor. The experimental stimuli where 4 video clips of the total duration of 6 seconds, in which either a human (a female; Figure 1A-D) or the robot (Figure 2A-D) gazed at a cup and then moved her/its right arm dropping the cup off the table. The size of the video stimuli is 1280×1024 pixels, while the AOIs were created of an equal size of 250x250 pixels.

Each video began with a scene in which the agent looked either at the camera or down at the table (2 s), thus defining the presence or absence of an initial eye-contact. Next, the agent turned toward and fixated on (2 s) the cup (congruent condition) or toward the portion of the table without the object (incongruent condition). In both conditions, the agent then moved her/its right arm (2 s) dropping the cup off the table. The human model maintained a neutral facial expression and re-

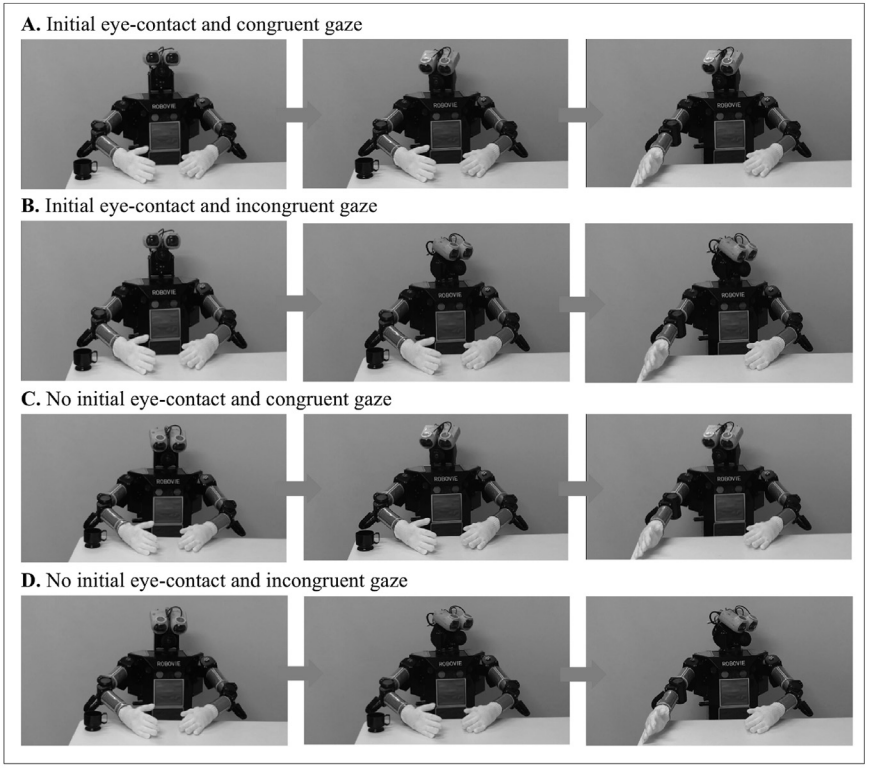
Figure 1 (A-D) - *Images representing the four human conditions*



mained silent throughout the entire sequence. Before each trial, an object (toy animation) appeared at the centre of the screen accompanied by a tinkling sound (2 s) to attract the infant's attention.

Before the test videos, infants were administered a familiarization trial that displayed the human or robot upper body (5 s) followed by the presentation of a fixation point paired with a beeping sound (1 s). The aim of this trial was to familiarize the infants with the setting and the agents (human or robot).

Figure 2 (A-D) - *Images representing the four robot conditions*



Procedure and apparatus

The infants were assessed individually in the presence of their mother in the Baby Centre at Kyoto University. Caregivers were instructed to close their eyes and to not talk to or interact with the infants during the task. The infants were randomly assigned to the Human or Robot con-

dition and the presentation order of each condition (presence/absence of initial eye-contact; congruent/incongruent gaze direction) was randomized across infants.

We used the Tobi T60 (Tobii pro studio, Tobii Technology, Stockholm) to record the infant's eye-gaze. The robot used for the tasks was the humanoid robot Robovie2 (Hiroshi Ishiguro Laboratories). The sampling rate of eye tracking was 60Hz. The participants were seated on the caregiver's lap approximately 60 centimeters from the monitor. Prior to recording, a five-point calibration was conducted. A clearview fixation filter was used for the eye-tracking data. Fixation was defined as gaze recorded within a 50-pixel diameter for a minimum of 200 ms, and this criterion was applied to the raw eye-tracking data to determine the duration of any fixation.

Data analysis

The infants had to complete at least 3 out of 4 trials (video stimuli) to be included in the final analysis. More specifically, if more than one trial produced no data on the dependent variables (due to the infant's disengagement or technical issues), the infant's data were completely removed from the analysis. For each video, 3 AOIs were defined as follows: 1) the agent's face, 2) the agent's hand and 3) the target object (the cup). Face and hand have been defined as areas of interest in line with recent studies that demonstrate their importance as early social cues (Fausey et al., 2016). In addition, each video was divided into three sequences: 1) an initial sequence in which there was no movement or action by the protagonist; 2) a sequence that included the movement of the gaze towards the object or its opposite position; 3) a final sequence representing the completion of the action, i.e. dropping the object from the table.

The dependent variables were the following: 1) the total fixation duration to evaluate infants' general attentional pattern on the different AOIs in the 3 sequences as a function of agency; 2) time to first fixation on the target in the final sequence showing the agent's action completion (anticipation), and assessing infants' sensitivity to the agents' referential gaze. Time to first fixation was calculated as the time interval between the beginning of sequence 3 (hand movement towards the target object) and the onset of the saccade from the infant's point of fixation at the end of the sequence 2 to the lateral target (in milliseconds). As outlined in the results, in sequence 2 infants fixated on the face area substantially more than on the target or the hand. Any possible differences in time to first fixation on the target in sequence 3 due to infants' fi-

nal gaze positioning in sequence 2 were counterbalanced by randomisation of conditions (initial eye contact/no eye contact, gaze congruence/gaze incongruence).

To evaluate infants' general attentional pattern to the video stimuli, total fixation duration was entered in a repeated measures GLM analysis with 2 levels of *eye-contact* (Present, Absent), 2 levels of *gaze-direction* (Congruent, Incongruent), 3 levels of *sequence* (No-Movement, Gaze-Direction, Hand-Movement) and 3 levels of *AOIs* (Face, Hand, Target) as within-subjects factors, and 2 levels of *agency* (Human, Robot) as the between-subjects factor. To assess the effect infants' sensitivity to the agents' referential gaze, time to first fixation on the target was entered in a repeated measures GLM analysis carried out only on sequence 3, with 2 levels of *eye-contact* (Present, Absent), 2 levels of *gaze-direction* (Congruent, Incongruent) as within-subjects factors, and 2 levels of *agency* (Human, Robot) as the between-subjects factor. The Greenhouse-Geisser correction was used for violations of Mauchly's Test of Sphericity ($p<.05$). *Post hoc* comparisons were Bonferroni corrected.

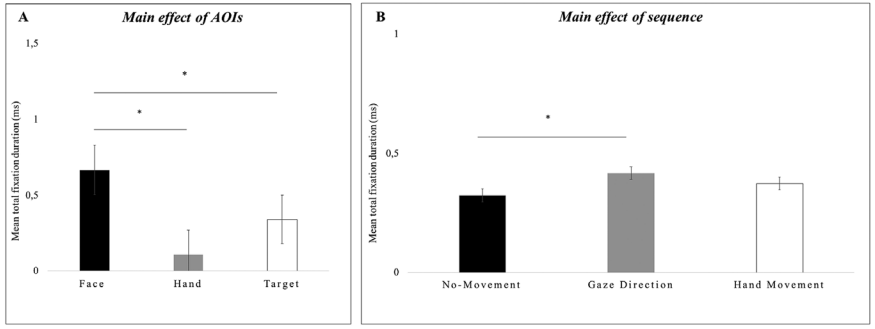
Results

Total fixation duration

The results showed a main effect of area of interests (*AOIs*), $F(2, 27) = 50.628$, $p<.001$, *partial-*$\eta^2=.789$, $\delta=1$, indicating that, independent of eye-contact, gaze direction, AOIs and agency, infants exhibited a greater fixation time on the face compared to both the hand, *M*diff=.557, *SE*=.05, $p<.001$ and the target, *M*diff=.326, *SE*=.07, $p<.001$. Additionally, infants' general attention was greater to the target compare to the hand, *M*diff=.231, *SE*=.048, $p<.001$ (Figure 3A). Additionally, the results revealed a main effect of *sequence*, $F(2, 56) = 5.187$, $p<.01$, *partial-*$\eta^2=.15$, $\delta=.80$, indicating that, independent of eye-contact, gaze direction, AOIs and agency, infants exhibited a shorter total fixation time in sequence 2 (Gaze-Direction), compared to sequence 1 (No-Movement), *M*diff=-.093, *SE*=.25, $p<.01$, but not with respect to sequence 3 (Hand-Movement). This main effect is plotted in Figure 3B. Overall, both results highlight the relevance, for infants, of the face region and gazing behaviour.

Furthermore, the results showed various interaction effects. A significant interaction between *AOI*eye-contact*, $F(2, 56) = 5.228$, $p<.01$, *partial-*$\eta^2=.15$, $\delta=.81$, revealed that the face was looked longer in the presence of eye-contact compared to the absence of eye-contact, *M*diff=.125, *SE*=.40, $p<.01$. Conversely, total fixation on the hand and target was not

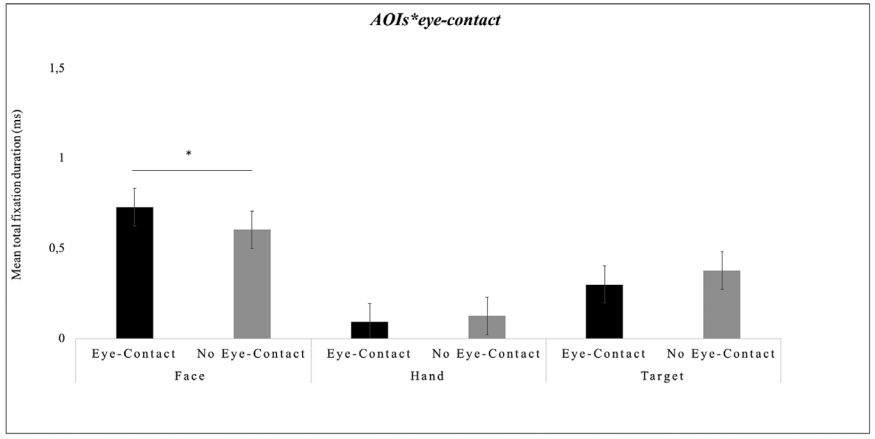
Figure 3 (A-B) - *Infants' total fixation duration scores*



Note: (A) The graph shows the mean scores (ms) for each area of interest (Face, Hand, Target) across eye-contact, gaze direction, video sequence and agent; (B) The graph shows the mean scores (ms) for each video sequence (No-Movement, Gaze-Direction, Hand-Movement) across areas eye-contact, gaze direction, area of interest and agent. The bars represent the standard error of the mean. * Indicates significant differences.

affected by the presence or absence of the initial eye-contact. This interaction is plotted in Figure 4.

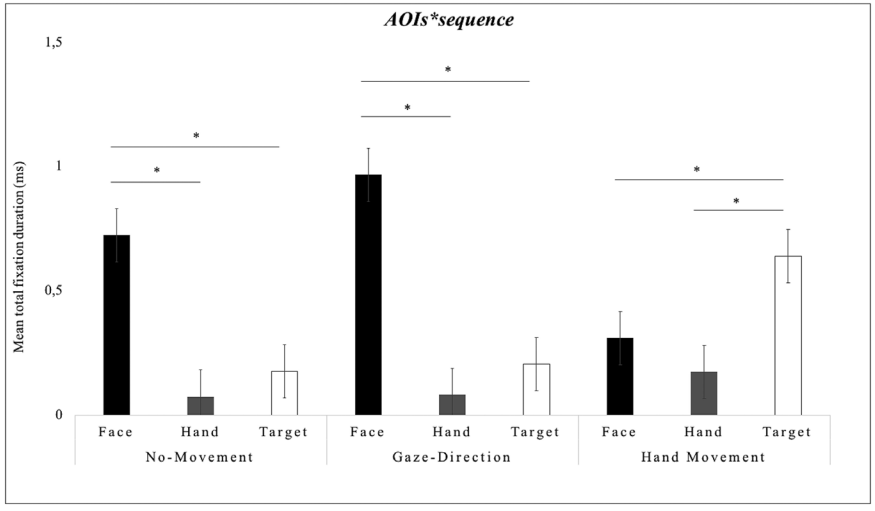
Figure 4 - *Infants' total fixation duration scores*



Note: The graph shows the mean scores (ms) as a function of area of interest (Face, Hand, Target) and as a function of eye-contact (Eye-Contact, No Eye-Contact). The bars represent the standard error of the mean. * Indicates significant differences.

Additionally, a significant two-way interaction *AOI*sequence*, $F(4, 112) = 48.396$, $p<.001$, *partial*-$\eta^2=.15$, $\delta=.81$, revealed that the face was looked at longer in sequence 2 (Gaze-Direction) compared to both sequence 1 (No-Movement), *M*diff=.243, *SE*=.078, $p<.05$, and sequence 3 (Hand-Movement), *M*diff=.658, *SE*=.078, $p<.001$. Furthermore, the *post hoc* analyses revealed that the hand was looked at longer in sequence 3 (Hand-Movement) compared to both sequence 1 (No-Movement), *M*diff=.100, *SE*=.028, $p<.01$, and sequence 2 (Gaze-Direction), *M*diff=.093, *SE*=.029, $p<.05$. Finally, the *post hoc* analyses revealed that the target was looked at longer in sequence 3 (Hand-Movement) compared to both sequence 1 (No Initial Movement), *M*diff=.463, *SE*=.08, $p<.001$, and sequence 2 (Hand-Movement), *M*diff=.435, *SE*=.052, $p<.001$.This interaction is plotted in Figure 5.

Figure 5 - *Infants' total fixation duration score*



Note: The graph shows the mean scores (ms) as a function of gaze (Congruent, Incongruent) and sequence (No-Movement, Gaze-Direction, Hand-Movement). The bars represent the standard error of the mean. * Indicates significant differences.

Finally, a significant three-way interaction (Table 1) *gaze-direction* sequence*agency*, $F(2, 24) = 3.241$, $p<.05$, *partial*-$\eta^2=.104$, $\delta=.595$, revealed, for the human condition only, greater total fixation duration on all AOIs in sequence 2 (Gaze-Direction) compared to sequence 1 (No-Movement), for both gaze conditions (Congruence, Incongruence).

Table 1 - *Statistics comparing total fixation duration between sequences (1. No-Movement; 2. Gaze-Direction; 3. Hand-Movement) as function of agency (Human, Robot), gaze (Congruent, Incongruent)*

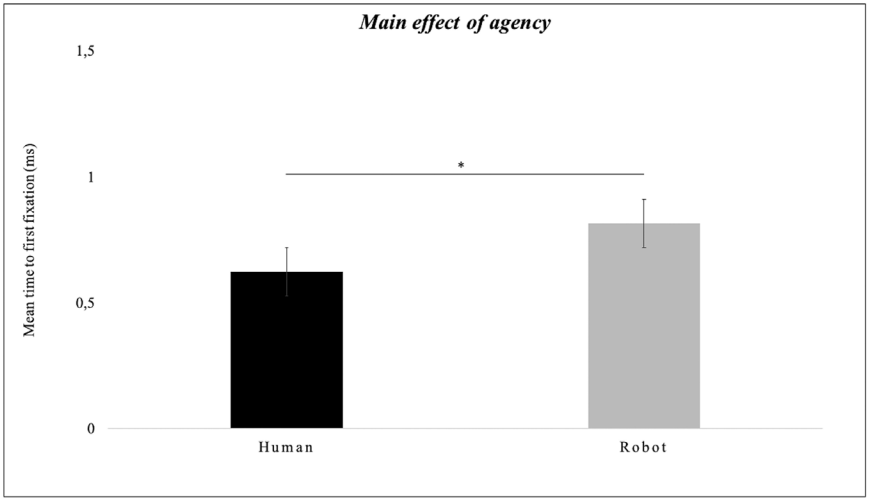
	Human					
	Gaze Congruence			*Gaze Incongruence*		
Sequence	*Mean Difference*	*Standard Error*	*Sign.*	*Mean Difference*	*Standard Error*	*Sign.*
1 vs 2	**-,144***	0,043	**0,007**	**-,101***	0,034	**0,018**
1 vs 3	0	0,068	1	-0,047	0,041	0,794
2 vs 3	0,144	0,06	0,069	0,054	0,033	0,354
	Robot					
	Gaze Congruence			*Gaze Incongruence*		
Sequence	*Mean Difference*	*Standard Error*	*Sign.*	*Mean Difference*	*Standard Error*	*Sign.*
1 vs 2	-0,052	0,046	0,796	-0,076	0,036	0,139
1 vs 3	-0,121	0,073	0,326	-0,031	0,044	1
2 vs 3	-0,069	0,064	0,88	0,044	0,036	0,674

* Correlation is significant at the level 0.05 (two-tailed).

Time to first fixation on the target

The results showed a main effect of *agency*, $F(1, 28) = 4.891$, $p<.05$, *partial*-η^2=.149, δ=.570, indicating that, independent of eye-contact and gaze direction, infants exhibited faster attendance to the target in the human condition compared to robot condition, *M*diff=-.192, *SE*=.087, $p<.05$.

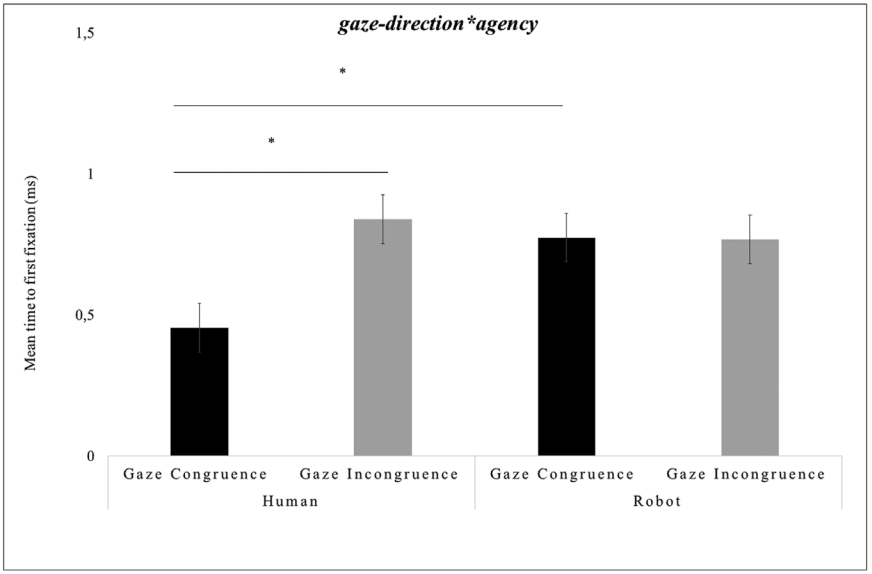
Figure 6 - *Infants' time to first fixation*



Note: The graph shows the mean scores (ms) as a function of agency (Human, Robot). The bars represent the standard error of the mean. * Indicates significant differences.

Additionally, the results showed a significant interaction between *gaze-direction*agency*, $F(2, 27) = 5.516$, $p<.01$, *partial*-η^2=.29, δ=.81. The *post hoc* analyses revealed that infants' attended faster to the target in the human gaze-congruent condition compared to human gaze-incongruent condition, *M*diff=-.411, *SE*=.138, $p<.005$. This result was not found in the robot condition. Finally, infants' time to first fixation on the target was faster in the human gaze-congruent condition compared to robot gaze-congruent condition, *M*diff=-.535, *SE*=.148, $p<.001$ (Figure 7).

Figure 7 - *Infants' time to first fixation scores*



Note: The graph shows the mean scores (ms) as a function of agent (Human, Robot) and gaze direction (Congruent, Incongruent). The bars represent the standard error of the mean. * Indicates significant differences.

Discussion

This study analysed 17-month-old infants' sensitivity to human and robot's referential gaze. To this purpose, the infants' gazing behaviour was assessed while watching videos showing either a human or a humanoid robot (Robovie) performing an action anticipated by either congruent or incongruent gaze toward a target. The results on total fixation duration showed that infants paid the greatest attention to the face area of both agents, and were equally engaged by both agents' initial eye-contact. However, infants showed to be particularly sensitive to the human gaze direction, as indexed by greater attention to head movement (compared to still face) in the human condition only. Sensitivity to the human gaze is also supported by data on time to first fixation, evaluating infants' response to the agents' referential gaze. The results showed that infants attended to the target faster when the action was performed by the human compared to the robot, and particularly in the presence of a congruent referential gaze.

With respect to the infants' attention pattern to the video stimuli,

indexed by total fixation time on specific AOIs (face, hand, object), the results showed that, independent of agent, infants preferred the face over both the hand and the target. Preference to the face was greater in the condition in which the action was primed by the agents' direct eye-contact toward the infant. These findings are in line with previous data on humans showing that the face attracts greater attention compared to other body parts already from 6 months of age (Blass & Camp, 2001; Marquis & Sugden, 2019). This effect is enhanced by the presence of direct eye-contact (Blass & Camp, 2001; Parise et al., 2008). Interestingly, our data enrich previous findings suggesting that, besides the face (Okumura et al., 2013a), also direct eye-contact of the robot exerts a similar attraction on the infant's attention as that observed for the human. These results suggest that the robot's anthropomorphic features and behaviours, such as the face and eye-contact, are able to activate similar attentional responses in infants aged 17 months. This might enable the robot to be ultimately perceived as a potential interactive partner. Nevertheless, we did find also differences in how the infant regards the referential nature of gaze between human and robot. First, our results on total fixation duration showed that infants paid greater attention to the human face in the sequence representing the gaze shift compared to the sequence in which there was not gaze shift. This effect was not found for the robot, suggesting that human referential gaze is more valuable (or informative) than the robot's gaze. Additionally, data on time to first fixation on the target showed both faster shift to the target in the human gaze, with respect to the robot gaze, congruent condition, and faster response to the target in the congruent vs. incongruent condition, for the human and not the robot. Infants' faster response to the congruent vs. incongruent human gaze shifts is in line with previous findings (Senju et al., 2006), also suggesting that this effect emerges as early as at nine months (Senju et al., 2006). As a counterproof, the differences between human and robot disappeared with gaze incongruence. These findings suggest that 17-months infants do not regard the robot's gaze as a social signal as they do for the human, where the inconsistency of gaze leads to a delayed response to the referential cue. Previous studies with 12-months infants (Okumura et al., 2013a) have shown that, despite gaze following was similar when attending to the human or robot's gaze, the human gaze was more informative and induced greater object learning compared to the robot's gaze. This suggests that infants may be keen to treat humans as prevalent sources of information. Also, and congruently with our findings, 12-month-olds, but not 10-months-old infants, showed greater anticipation time on the target object when primed by the human than the robot's gaze (Okumura et al., 2013a), thus sup-

porting that infants privilege the referential nature of the human gaze compared to the robot's gaze.

The mechanism of referential gaze recognition at 17 months is strictly related to human social cognition. The preferential attitude exhibited by infants for the referential gaze of a conspecifics compared to the gaze of a robot leads to at least two possible interpretations. The first one accounts for the specificity of a socio-cognitive mechanism, which would not allow the generalisation to other species or entities (e.g., robots). This would be in line with the hypothesis of a human species-specific eye-gaze mechanism suggested by studies demonstrating that infants prefer human gaze compared to the gaze of other living species, like monkeys (Kano & Call, 2014), or non-living ones, like robots (Meltzoff et al., 2010; Okumura et al., 2013a). The second interpretation would take on the experiential account and argue that preference for the human gaze is due to the inexperience that infants have with the gaze of other species or entities, like robots. Several studies have shown that gaze following develops within 12 months (Butterworth & Itakura, 2000; Brooks & Meltzoff, 2005; Carpenter et al., 1998; Deák et al., 2000; Hofsten et al., 2005; Moore & Corkum, 1994; Mundy & Newell, 2007), although the debate is still open on the exact onset period (Astor & Gredebäck, 2019; D'Entremont et al., 1997; Farroni et al., 2004; Hood et al., 1998; Imafuku et al., 2017; Ishikawa & Itakura, 2019; Senju & Csibra, 2008). At this age, infants appear to be selectively attuned to follow the human gaze, and this would be in line with findings showing that 12-month-old infants have difficulties in following the gaze of other species (Kano & Call, 2014). This selectivity may be plausibly due to extensive exposure to human interaction in early life. Generalization of gaze-following may then occur in subsequent stages of development, possibly through experience, i.e. learning. The importance of learning is supported by evidence suggesting that the mechanism of gaze-following gradually emerges through social experience (Deák et al., 2014; Ishikawa et al., 2020; Triesch et al., 2006). Accordingly, infants would display preference for the human referential gaze because had not had the experience of the robot's gaze. The hypothesis of a narrowing of such inborn gaze sensitivity through extensive experience with human gaze is also supported by studies with precocial bird species suggesting that the mechanism for gaze following is highly conserved among vertebrates, is active at birth, and therefore it is not species specific (Jaime et al., 2009; Rosa Salva et al., 2007). Future studies could specifically address this interpretation by, for example, exposing infants of different ages to several sessions with robots and observe if sensitivity to referential gaze increases with experience. Also, robots could be placed in ecological settings for infants where they typically experience human interaction.

In sum, our data indicate, on the one hand, that children recognise the physical traits and behaviours that are salient for social interaction, such as the face and direct eye-contact, in both agents. Conversely, when agents perform an action – look at an object –, infants appear to be more sensitive to the human than the robot agency. This data leads to an important consideration regarding the possible presence of two distinct levels of gaze-following mechanism: a first preparatory level that focuses the infant's attention on the parts considered salient for interaction, allowing to recognise the other as a possible interactive partner; and a second level that evaluates whether the partner has agency. The results of the present study seem to suggest that at 17 months the first level of the gaze-following mechanism – recognise the other as a plausible interactional partner – is both active and generalised, while the second level – attribution of agency to the partner – is active with human but not (yet) generalised to other species or entities, i.e. robots.

Conclusions

Infants at 17 months of age do not recognize the referential nature of the robot's gaze: this type of attribution, referential, is more readily made with conspecifics. Nevertheless, infants seem to be sensitive to some of the robot's physical and behavioural human attributes, and namely its face and eye-contact. More specifically, data showed that children recognize, on the one hand, the robot as a plausible interactive partner but, at the same time, do not ascribe agency to it. These results allowed us to hypothesize two distinct processing levels of the gaze-following mechanism: the first one that evaluates if an entity is an interactive partner; the second one that evaluates if it is endowed with agency. Infants at 17 months of age seem to be particularly sensitive to human features also when embedded in a robot (first processing level), but they are not equipped with a generalization mechanism that enables them to regard the robot as a human-like agent (second processing level). As suggested, this mechanism may develop with experience.

Limitations and Future Directions

Since at this age infants do not seem to react to the referential nature of the robot's gaze, it might be important to study whether these effects change as function of age. Additionally, it is important to underline that in our study the infants might have followed the head movement rather than just the eye-gaze. This choice was necessary because

it was not possible to recreate in the robot the contrast between the dark pupil and the white sclera that guides eye-gaze perception. Therefore, future studies should find the way to test these two effects (head movement and eye-gaze) independently. Also, the results presented in this study were acquired with only one type of robot, while it might be interesting and important to evaluate the effect of robots differing in physical features, with different cues of embodiment/anthropomorphization. Moreover, it would be interesting to specifically compare in future related studies two alternative interpretations of the data: the idea of a species-specific gaze-following mechanism that is not generalized to robots vs. the lack of experience with robots. Finally, in line with current theoretical claims that studies with robots can be used to understand the nature of human psychological mechanisms (Wiese et al., 2017; Wykowska et al., 2016), this line of research with infants and robots can shed light on the mechanisms underpinning the development of early social cognition.

References

Astor, K., & Gredebäck, G. (2019). Gaze following in 4.5- and 6-month-old infants: The impact of proximity on standard gaze following performance tests. *Infancy, 24*(1), 79-89.

Biondi, M., Boas, D. A., & Wilcox, T. (2016). On the other hand: Increased cortical activation to human versus mechanical hands in infants. *NeuroImage, 141*, 143-153.

Blass, E. M., & Camp, C. A. (2001). The ontogeny of face recognition: Eye contact and sweet taste induce face preference in 9- and 12-week-old human infants. *Developmental Psychology, 37*(6), 762-774.

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science, 8*(6), 535-543.

Butterworth, G., & Cochran, E. (1980). Towards a mechanism of joint visual attention in human infancy. *International Journal of Behavioral Development, 3*(3), 253-272.

Butterworth, G., & Itakura, S. (2000). How the eyes, head and hand serve definite reference. *British Journal of Developmental Psychology, 18*(1), 25-50.

Butterworth, G., & Jarrett, N. (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology, 9*(1), 55-72.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(4), i.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148-153.

Deák, G. O., Flom, R. A., & Pick, A. D. (2000). Effects of gesture and target on 12- and 18-month-olds' joint visual attention to objects in front of or behind them. *Developmental Psychology, 36*(4), 511-523.

Deák, G. O., Krasno, A. M., Triesch, J., Lewis, J., & Sepeta, L. (2014). Watch the hands: Infants can learn to follow gaze by seeing adults manipulate objects. *Developmental Science, 17*(2), 270-281.

D'Entremont, B., Hains, S. M. J., & Muir, D. W. (1997). A demonstration of gaze following in 3- to 6-month-olds. *Infant Behavior and Development, 20*(4), 569-572.

Di Dio, C., Manzi, F., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., & Marchetti, A. (2020). It does not matter who you are: Fairness in pre-schoolers interacting with human and robotic partners. *International Journal of Social Robotics, 12*(5), 1045-1059.

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., & Marchetti, A. (2020b). Shall I trust you? from child-robot interaction to trusting relationships. *Frontiers in Psychology, 11*, 469.

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., & Marchetti, A. (2020c). Come i bambini pensano alla mente del robot: il ruolo dell'attaccamento e della Teoria della Mente nell'attribuzione di stati mentali ad un agente robotico [How children think about the robot's mind. The role of attachment and Theory of Mind in the attribution of mental states to a robotic agent]. *Sistemi Intelligenti, 1*(20), 41-56.

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition, 6*(5), 509-540.

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences, 99*(14), 9602-9605.

Farroni, T., Massaccesi, S., Pividori, D., & Johnson, M. H. (2004). Gaze following in newborns. *Infancy, 5*(1), 39-60.

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition, 152*, 101-107.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review, 5*(3), 490-495.

Gredebäck, G., Fikke, L., & Melinder, A. (2010). The development of joint visual attention: A longitudinal study of gaze following during interactions with mothers and strangers: The development of joint visual attention. *Developmental Science, 13*(6), 839-848.

Gredebäck, G., Theuring, C., Hauf, P., & Kenward, B. (2008). The microstructure of infants' gaze as they view adult shifts in overt attention. *Infancy, 13*(5), 533-543.

Hofsten, C., Dahlström, E., & Fredriksson, Y. (2005). 12-month-old infants' perception of attention direction in static video images. *Infancy, 8*(3), 217-231.

Hood, B. M., Willen, J. D., & Driver, J. (1998). Adult's eyes trigger shifts of visual attention in human infants. *Psychological Science, 9*(2), 131-134.

Imafuku, M., Kawai, M., Niwa, F., Shinya, Y., Inagawa, M., & Myowa-Yamakoshi, M. (2017). Preference for dynamic human images and gaze-following abilities in preterm infants at 6 and 12 months of age: An eye-tracking study. *Infancy, 22*(2), 223-239.

Ishikawa, M., & Itakura, S. (2018). Observing others' gaze direction affects infants' preference for looking at gazing- or gazed-at faces. *Frontiers in Psychology, 9*, 1503.

Ishikawa, M., & Itakura, S. (2019). Physiological arousal predicts gaze following in infants. *Proceedings of the Royal Society B: Biological Sciences, 286*(1896), 20182746.

Ishikawa, M., Senju, A., & Itakura, S. (2020). Learning process of gaze following: Computational modeling based on reinforcement learning. *Frontiers in Psychology, 11*, 213.

Ishikawa, M., Yoshimura, M., Sato, H., & Itakura, S. (2019). Effects of attentional behaviours on infant visual preferences and object choice. *Cognitive Processing, 20*(3), 317-324.

Itakura, S., Ishida, H., Kanda, T., Shimada, Y., Ishiguro, H., & Lee, K. (2008). How to build an intentional android: Infants' imitation of a robot's goal-directed actions. *Infancy, 13*(5), 519-532.

Jaime, M., Lopez, J. P., & Lickliter, R. (2009). Bobwhite quail (Colinus virginianus) hatchlings track the direction of human gaze. *Animal Cognition, 12*(4), 559-565.

Kanakogi, Y., & Itakura, S. (2011). Developmental correspondence between action prediction and motor ability in early infancy. *Nature Communications, 2*(1), 341.

Kano, F., & Call, J. (2014). Cross-species variation in gaze following and conspecific preference among great apes, human infants and adults. *Animal Behaviour, 91*, 137-150.

Langdon, R., & Smith, P. (2005). Spatial cueing by social versus nonsocial directional signals. *Visual Cognition, 12*(8), 1497-1527.

Marchetti, A., Manzi, F., Itakura, S., & Massaro, D. (2018). Theory of Mind and humanoid robots from a lifespan perspective. *Zeitschrift Für Psychologie, 226*(2), 98-109.

Marquis, A. R., & Sugden, N. A. (2019). Meta-analytic review of infants' preferential attention to familiar and unfamiliar face types based on gender and race. *Developmental Review, 53*, 100868.

Meltzoff, A. N., Brooks, R., Shon, A. P., & Rao, R. P. N. (2010). "Social" robots are psychological agents for infants: A test of gaze following. *Neural Networks, 23*(8), 966-972.

Moore, C., & Corkum, V. (1994). Social understanding at the end of the first year of life. *Developmental Review, 14*(4), 349-372.

Morales, M., Mundy, P., & Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behavior and Development, 21*(2), 373-377.

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current Directions in Psychological Science, 16*(5), 269-274.

Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., & Itakura, S. (2013a). The power of human gaze on infant learning. *Cognition, 128*(2), 127-133.

Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., & Itakura, S. (2013b). Infants understand the referential nature of human gaze but not robot gaze. *Journal of Experimental Child Psychology, 116*(1), 86-95.

Parise, E., Reid, V. M., Stets, M., & Striano, T. (2008). Direct eye contact influences the neural processing of objects in 5-month-old infants. *Social Neuroscience, 3*(2), 141-150.

Phillips, A. T., Wellman, H. M., & Spelke, E. S. (2002). Infants' ability to connect gaze and emotional expression to intentional action. *Cognition, 85*(1), 53-78.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*(1), 3-25.

Rosa Salva, O., Regolin, L., & Vallortigara, G. (2007). Chicks discriminate human gaze with their right hemisphere. *Behavioural Brain Research, 177*(1), 15-21.

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology, 18*(9), 668-671.

Senju, A., & Johnson, M. H. (2009). The eye contact effect: Mechanisms and development. *Trends in Cognitive Sciences, 13*(3), 127-134.

Senju, A., Johnson, M. H., & Csibra, G. (2006). The development and neural basis of referential gaze perception. *Social Neuroscience, 1*(3-4), 220-234.

Slaughter, V., & McConnell, D. (2003). Emergence of joint attention: Relationships between gaze following, social referencing, imitation, and naming in infancy. *The Journal of Genetic Psychology, 164*(1), 54-71.

Triesch, J., Teuscher, C., Deak, G. O., & Carlson, E. (2006). Gaze following: Why (not) learn it? *Developmental Science, 9*(2), 125-147.

Wiese, E., Metta, G., & Wykowska, A. (2017). Robots as intentional agents: Using neuroscientific methods to make robots appear more social. *Frontiers in Psychology, 8*, 1663.

Woodward, A. L. (2003). Infants' developing understanding of the link between looker and object. *Developmental Science, 6*(3), 297-311.

Wykowska, A., Chaminade, T., & Cheng, G. (2016). Embodied artificial agents for understanding human social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1693), 20150375.

5. The Robot Made Me Do It

Human-Robot Interaction and Risk-Taking Behavior

Y. Hanoch, F. Arvizzigno, D. Hernandez García, S. Denham, T. Belpaeme, M. Gummerum

ABSTRACT

Empirical evidence has shown that peer pressure can impact human risk-taking behavior. With robots becoming ever more present in a range of human settings, it is crucial to examine whether robots can have a similar impact. Using the balloon analogue risk task (BART), participants' risk-taking behavior was measured when alone, in the presence of a silent robot, or in the presence of a robot that actively encouraged risk-taking behavior. In the BART, shown to be a proxy for real risk-taking behavior, participants must weigh risk against potential payout. Our results reveal that participants who were encouraged by the robot did take more risks, while the mere presence of the robot in the robot control condition did not entice participants to show more risk-taking behavior. Our results point to both possible benefits and perils that robots might pose to human decision-making. Although increasing risk-taking behavior in some cases has obvious advantages, it could also have detrimental consequences that are only now starting to emerge.

Introduction

Can robots influence and change humans' behavior? This study addressed this question by focusing on whether robots can alter human risk-taking behavior. Risk taking is a key human behavior that has major financial, health, and social implications and has been shown to be subject to the influence of others. Gaining insights into whether robots af-

---

This chapter was originally published as Hanoch, Y., Arvizzigno, F., Hernandez García, D., Denham, S., Belpaeme, T., & Gummerum, M. (2021). The robot made me do it: Human-robot interaction and risk-taking behavior. *Cyberpsychology, Behavior, and Social Networking, 24*(5), 337-342. The authors have no commercial associations that might create a conflict of interest in connection with the submitted articles. None of the authors has any competing financial interests. No funding was received. Supplementary Data: Supplementary Table S1.

fect human risk-taking behavior thus has clear ethical, policy, and theoretical implications.

One area of research has explored robots' ability to exert peer pressure, more specifically, whether people follow the incorrect judgments and behaviors of robots. Drawing on Asch's (1951) classic work – showing that individuals conform to a unanimous majority's incorrect judgments – studies (Brandstetter et al., 2014; Salomons et al., 2018; Vollmer et al., 2018; Xu & Lombard, 2017) have examined whether humans would conform to a unanimous but incorrect group of robots. One investigation demonstrated that participants showed conformity when interacting with human peers, but not with robots. Other studies (Salomons et al., 2018; Vollmer et al., 2018) reported that human participants did show conformity when interacting with robot peers or that adults resisted robot peer pressure, but young children conformed.

Peer pressure from other humans also plays a significant role in individuals' risk-taking behavior. For example, researchers (Gardner & Steinberg, 2005) examined whether the mere presence of peers impacted risk-taking behavior in participants. Participants who completed a self-report questionnaire and a behavioral risk-taking task in the presence of peers focused more on the benefits compared with the risks and, importantly, exhibited riskier behavior. Focusing on peer pressure in risky driving, the leading cause of death among young adults (Steinberg & Monahan, 2007; World Health Organization, 2018), in two studies (Shepherd et al., 2011), university students were placed in a driving simulator either by themselves or with confederate peers posing as passengers. The confederates' role was to encourage the drivers to engage in riskier driving behavior. In line with the researchers' prediction, the confederates' encouragement led to riskier behavior (e.g., driving faster) and higher accident rates (Gheorghiu et al., 2015; Pradhan et al., 2014; Simons-Morton et al., 2005; Toxopeus et al., 2011).

Whether the effect of peer pressure on risk taking would emerge in interactions with robots is an open, and important, question. Given the paucity of previous research coupled with methodological and ethical issues, it is impossible at this stage to know whether robots could increase risky behaviors such as smoking and substance abuse. However, we can use a risk-taking measure that has been linked to real-life risky behavior and has been shown to be impacted by the presence of a peer. One such measure is the balloon analogue risk task (BART) in this study (Lejuez et al., 2002; Hopko et al., 2006; Lejuez et al., 2003, 2004; Reynolds et al., 2014).

The present study was designed to examine whether robots would impact participants' risk-taking behavior. Following earlier work with humans (Pradhan et al., 2014; Simons-Morton et al., 2005; Toxopeus

et al., 2011; Reynolds et al., 2014), participants completed the BART alone (control condition), in the mere presence of a silent robot that did not interact with or encourage any risky behavior from the participant (robot control condition), or in the presence of a robot that interacted with the participants and provided explicit statements encouraging risk-taking (experimental condition). It was predicted that participants who completed the BART in the experimental condition (risk-encouraging robot) would exhibit higher risk-taking behavior compared with the two control groups. Because previous research (Gardner & Steinberg, 2005) has shown that the mere presence of a human peer facilitates risk-taking, we also examined whether the presence of a silent noninteractive robot (robot control condition) would have a similar effect.

Materials and Methods

Participants

Ethics approval was granted before the commencement of the study. A total of 180 undergraduate psychology students participated in the study (154 women, 26 men; M_{age} = 21.43 years, SD = 7). Participants were randomly allocated to one of three conditions: control (N = 60, 50 females, 10 males), robot control (N = 60, 54 females, 6 males), and experimental (N = 60, 50 females, 10 males). One female participant from the experimental condition was removed from the analyses because of malfunctioning equipment; therefore, the experimental condition contained N = 59 participants (49 females, 10 males). Participants in the three conditions did not differ in age, $F(2, 178) = 0.18$, $p = 0.84$, or sex, χ^2 (4) = 3.31, $p = 0.51$. Participants received course credit and financial earnings (1 U.K. penny for each pump) on the BART.

Materials

Balloon analogue risk task. Over 30 trials, participants were asked to press the space bar on a computer keyboard to inflate a balloon displayed on the computer monitor (Lejuez et al., 2002). In total, thus, participants inflated 30 different balloons. With each press of the space bar, the balloon was inflated by 1°, and 1 cent (U.K. currency) was added to the participant's 'temporary money bank,' which was shown on the screen. This represented the sum earnings for the current balloon. After each pump, a 'Collect reward' button displayed on screen could be

clicked by the participant to 'cash in' the winnings for the current balloon. By clicking the button, the participant moved on immediately to the next balloon and the winnings for the previous balloon were added to the participant's overall earnings, also displayed on screen. If, however, the balloon exploded after a pump was made, all winnings for that balloon were lost and participants moved on to the next balloon without adding to their overall earnings. A random number generator determined at when the balloon would explode, with the constraint that the probability that a balloon would explode increases with each pump that was made (1/128, 1/127, etc.). The highest number of possible pumps was 128. Each participant received a unique series of balloon explosion points for the 30 balloons/trials.

For each balloon, the following scores were derived: (1) the number of pumps made by participants; (2) the explosion point of each balloon (randomly determined by the program - see above); (3) whether the balloon exploded or not; and (4) participants' earnings (in U.K. pennies) for each balloon. Number of pumps, explosions, and earnings were summed up across the 30 trials.

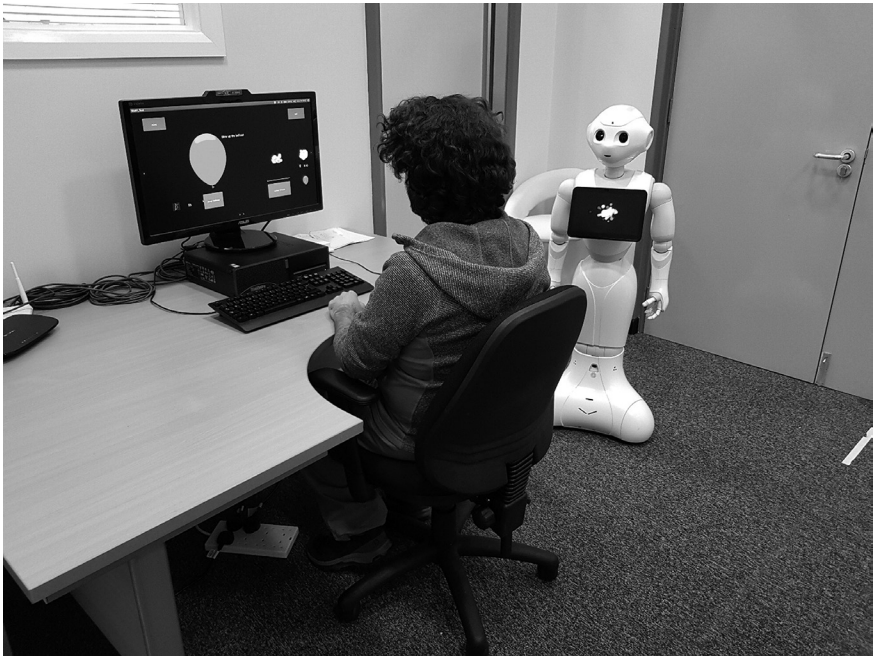
Godspeed. The Godspeed measures participants' attitudes toward robots on five subscales, anthropomorphism (5 items; $\alpha = 0.82$), animacy (6 items; $\alpha = 0.85$), likeability (5 items; $\alpha = 0.91$), perceived intelligence (5 items; $\alpha = 0.82$), and perceived safety (6 items; $\alpha = 0.70$), with items rated on a 5-point semantic differential rating scale (Bartneck et al., 2009). Due to the strong positive and significant correlations between all subscales, r's$(179) = 0.33$-0.72, all p's< 0.001, scores were averaged to create one 'robot impression' score; higher scores represent more positive impressions of the robot.

Self-reported risk-taking. Participants' self-reported risk-taking attitude was measured by a single item (Dohmen et al., 2011): 'How do you see yourself? Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?'. Participants were asked to indicate on a Likert-type scale of 0 (*not at all willing to take risks*) to 7 (*very willing to take risks*) how willing they are to take risks.

Robot. One SoftBank Robotics Pepper robot was used in the two robot conditions (Figure 1). Pepper, 1.21-meter tall with 25 degrees of freedom, is a medium-sized humanoid robot designed primarily for human-robot interaction (HRI). The robot was fully autonomous, running bespoke software that allowed it to be controlled by the software running on the experimenter's laptop. This robot performed scripted behaviors that were identical for all participants in a condition (see the

Additional Experimental Materials[1] section in the Supplementary Data[2]). The robot stood on the floor beside the participants' seating arrangement.

Figure 1 - *Overview of the experimental setup and visual stimulus*



Methods

All participants completed the experiment in the same laboratory room (Figure 1). The control condition participants completed the study in the laboratory and were provided with the same general instructions as the two experimental groups, using the computer screen only. The robot control condition participants completed the study in the same laboratory, but in this case, Pepper the robot was present in the room and provided participants with *only* the study instructions. For participants in the experimental condition, the robot provided instructions and, importantly, encouraging statements (e.g., 'Why did you stop pumping?').

[1] https://www.liebertpub.com/doi/10.1089/cyber.2020.0148#s004.

[2] https://www.liebertpub.com/doi/suppl/10.1089/cyber.2020.0148/suppl_file/Supp_Data.docx.

Encouragements by the robot were given during the experiment both in cases where participants stopped pumping before they reached 50 pumps and in cases where the balloon exploded (see Supplementary Data[3]). The robot used one of the statements in random order.

After participants completed the BART, they were asked to complete two manipulation checks: the single-item self-assessment of their risk-taking, followed by the Godspeed questionnaire. We decided to administer the Godspeed in all three conditions to make the participants' experience of the study maximally comparable. At the end of the study, participants were paid their earnings, thanked, and debriefed verbally and in writing.

Results

Manipulation check

A one-way analysis of variance (ANOVA) revealed significant differences in how participants in the three conditions perceived the robot, $F(2, 176) = 14.28$, $p < 0.001$. *Post hoc* tests (with Bonferroni corrections) indicated that participants in the control condition had a significantly lower positive impression of the robot than participants in the robot control or the experimental condition (all p's < 0.001) (Table 1). Impressions of the robot did not differ between participants in the experimental and robot control conditions ($p = 1.00$; see Supplementary Data[4] and Supplementary Table S1[5] for analyses of the Godspeed subscales).

Table 1 - *Manipulation check: means (and standard deviations) of self-reported risk taking and robot impression by condition*

Variable	*Control group (N = 60)*	*Robot control group (N = 60)*	*Experimental group (N = 59)*
Self-reported risk-taking	*M* = 5.38	*M* = 4.62	*M* = 4.92
	SD = 1.72	*SD* = 1.53	*SD* = 1.70
Robot impression	*M* = 2.89	*M* = 3.34	*M* = 3.40
	SD = 0.62	*SD* = 0.52	*SD* = 0.55

[3] https://www.liebertpub.com/doi/suppl/10.1089/cyber.2020.0148/suppl_file/Supp_Data.docx.

[4] https://www.liebertpub.com/doi/suppl/10.1089/cyber.2020.0148/suppl_file/Supp_Data.docx.

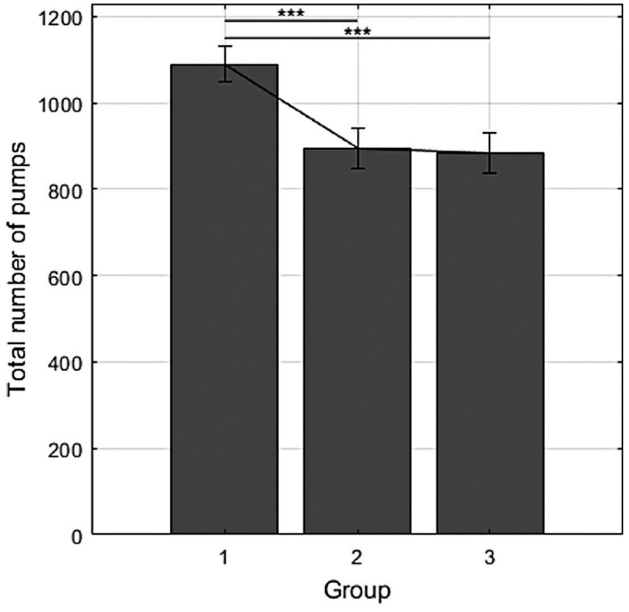
[5] https://www.liebertpub.com/doi/suppl/10.1089/cyber.2020.0148/suppl_file/Supp_Data.docx.

A one-way ANOVA also showed a significant difference in participants' self-assessment of their own risk-taking tendencies, $F(2, 176) = 3.29$, $p = 0.04$. Those in the control condition indicated significantly higher risk-taking tendencies than those in the robot control condition ($p = 0.04$), but did not differ from participants in the experimental condition ($p = 0.37$). Risk-taking tendencies of participants in the robot control and the experimental conditions did not differ ($p = 0.98$) (Table 1).

Risk taking

A Poisson regression indicated a significant effect of condition on the number of pumps, $\chi^2(2) = 713.09$, $p < 0.001$. The number of pumps across the 30 rounds was significantly higher in the experimental condition than in the control condition, $B = -0.17$, $SE = 0.01$, Wald $\chi^2(1) = 559.17$, $p < 0.001$, and the robot control condition, $B = -.15$, $SE = 0.01$, Wald $\chi^2(1) = 482.63$, $p < 0.001$. The median number of pumps in the experimental condition was 1.23 times higher than in the control condition and 1.22 times higher than in the robot control condition (Figure 2).

Figure 2 - *Total number of pumps*

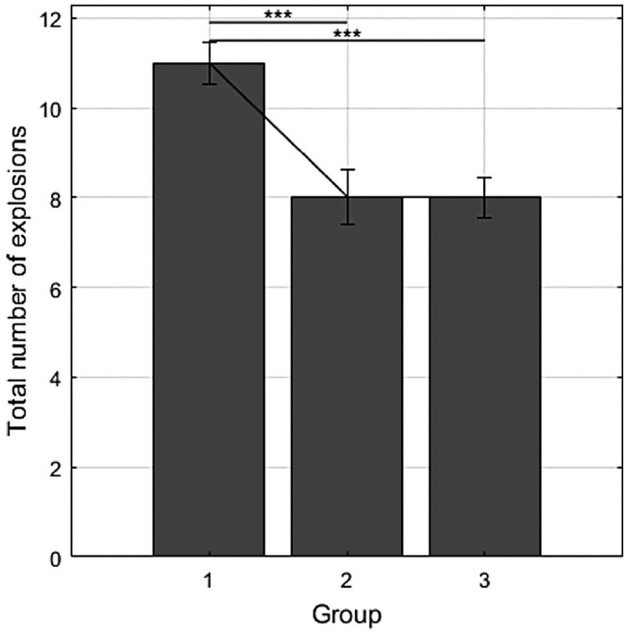


Note: Bars show the median total number of pumps for each group. Whiskers indicate standard error. Group 1: Experimental condition; Group 2: Robot control condition; Group 3: Control condition (***$p < 0.001$).

Spearman correlations indicated that there was no significant relationship between number of pumps and self-reported risk-taking tendencies, $\rho(178) = 0.06$, $p = 0.42$.

A significant effect also emerged for the number of explosions, $\chi^2(2) = 30.46$, $p < 0.001$. Participants experienced more explosions in the experimental than in the control condition, $B = -0.32$, $SE = 0.06$, Wald $\chi^2(1) = 27.61$, $p < 0.001$, and the robot control condition, $B = -.23$, $SE = 0.06$, Wald $\chi^2(1) = 14.50$, $p < 0.001$. The median number of explosions was 1.38 times higher in the experimental than in the control condition and 1.38 times higher than in the robot control condition (Figure 3). The number of explosions did not significantly correlate with self-reported risk-taking tendencies, $\rho(178) = 0.11$, $p = 0.09$.

Figure 3 - *Total number of explosions*

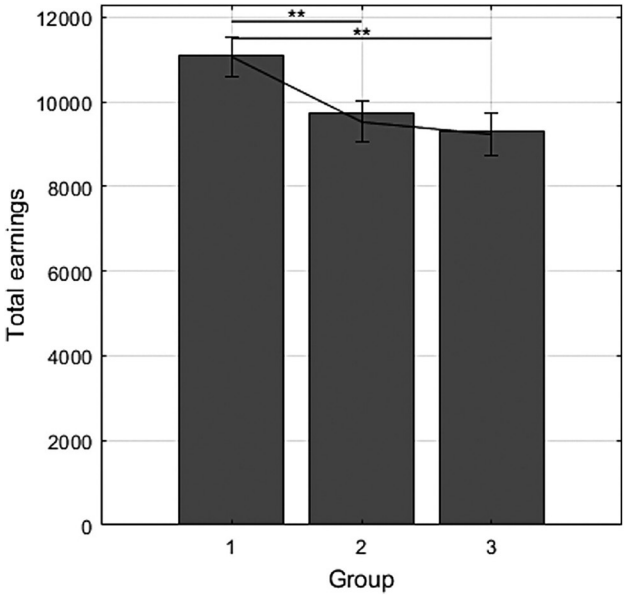


Note: Bars show the median total number of explosions for each group. Whiskers indicate standard error. Group 1: Experimental condition; Group 2: Robot control condition; Group 3: Control condition (***$p < 0.001$).

Participants in the experimental condition also earned significantly more, on average, than those in the control condition ($p = 0.02$) and the robot control condition ($p = 0.03$), $F(2, 176) = 4.70$, $p = 0.01$ (Figure

4). Participants in the experimental condition earned on average 1.20 times more than those in the control condition and 1.16 times more than those in the robot control condition. Earnings did not significantly correlate with self-reported risk-taking tendencies, $r(178) = 0.08$, $p = 0.31$.

Figure 4 - *Total earnings*

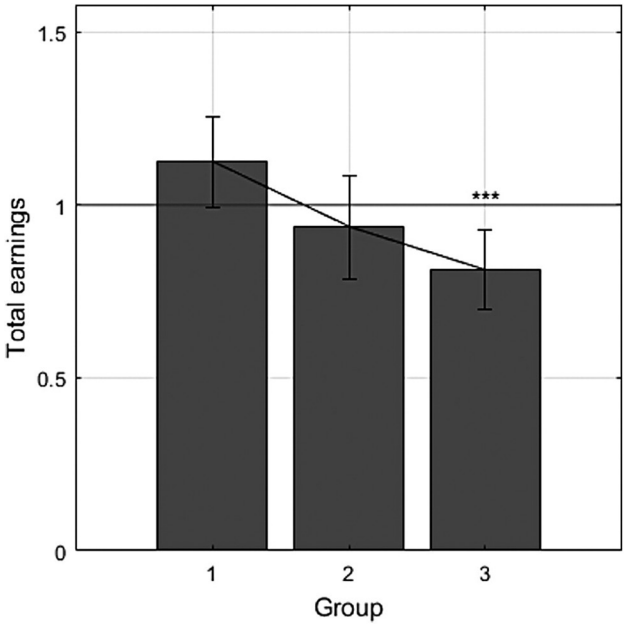


Note: Bars show the median total earnings for each group. Whiskers indicate standard error. Group 1: Experimental condition; Group 2: Robot control condition; Group 3: Control condition (**$p < 0.01$).

Why did participants in the experimental condition earn more than those in the control conditions despite experiencing more explosions, which wiped out their earnings in any round in which the balloon exploded? Participants in the experimental condition tended not to reduce their number of pumps in response to an explosion. Below we quantify this 'explosion effect' as the number of pumps in the trial after an explosion divided by the number of pumps in the trial before an explosion. An explosion effect <1 indicates a reduction in pumps after an explosion; an explosion effect >1 indicates an increase in pumps after an explosion. The median explosion effects were 1.13 in the experimental, 0.94 in the robot control, and 0.81 in the control condition (Figure 5). Binomial tests indicated that the median explosion effect did not

differ from 1 in the experimental (p=0.26) and robot control (p=0.12) conditions, but was significantly smaller than 1 in the control condition (p=0.007). Thus, participants in the control condition reduced their pumps after experiencing an explosion. A Kruskal-Wallis H test showed that the medians of the three conditions significantly differed from each other, χ^2 (2) = 11.01, p=0.004.

Figure 5 - *Explosion effect on subsequent number of pumps*



Note: Bars show the median explosion effect for each group. The explosion effect quantifies how much experiencing an explosion influences subsequent behavior, and is calculated as the number of pumps after an explosion divided by number of pumps before the explosion. An explosion effect <1 indicates a reduction in pumps after an explosion; an explosion effect >1 indicates an increase in pumps after an explosion; an explosion effect of 1 (horizontal line) indicates no change. Whiskers indicate standard error. Group 1: Experimental condition; Group 2: Robot control condition; Group 3: Control condition (***p<0.001).

Discussion

Can robots exert peer pressure to impact human risk-taking behavior? Our results reveal that participants who were encouraged by the ro-

bot did indeed take more risks. They pumped the balloon significantly more often, experienced a higher number of explosions, and earned significantly more money. Thus, our results suggest that the robot's encouragement to take additional risks seemed to have influenced participants' risk-taking behavior in the BART.

It is notable that the mere presence of the robot in the robot control condition did not entice participants to show more risk-taking behavior. In fact, on the three indices of risk taking measured by the BART (i.e., number of pumps, number of explosions, average earnings), participants in the robot control condition behaved strikingly like those in the control condition. These differences in risk taking between the experimental and robot control conditions cannot be explained by self-reported risk-taking tendencies, because those did not differ between the two groups. Similarly, participants in the experimental and robot control conditions did not differ in their impressions of the robot. These findings therefore contrast with studies (Chou & Nordgren, 2017; Gardner & Steinberg, 2005) showing that the mere presence of human peers increases risk taking. Evaluation apprehension, people's concern that they might be negatively evaluated by others, has been proposed as one of the explanations as to why the mere presence of human peers facilitates changes in behaviors (Cottrell et al., 1968; Guerin & Innes, 1984). As such, in the robot control condition, participants might not have perceived the silent noninteractive robot as evaluating them, and thus, the mere presence of the robot did not impact their risk taking. While previous research (Gray et al., 2007; Terada et al., 2007) has shown that humans can attribute mental states (e.g., intentions, agency) to robots and other inanimate objects, this process is not automatic but might depend on robots being active and interacting with the participant. Future research might explore further whether the mere presence of a robot in facilitating risk taking is indeed based on evaluation apprehension and the attribution of a mental states to the robot.

Our study not only reveals differences in risk-taking by condition but also suggests a possible mechanism underlying these differences. Specifically, results on the 'explosion effect' indicate that participants in the control condition seemed to learn from the negative experiences by reducing their risk taking (i.e., number of pumps) after they experienced an explosion. In contrast, experiencing an explosion did not alter the risk-taking behavior of participants in the experimental and robot control conditions. In other words, while participants in the control condition scaled back their risk-taking behavior following a balloon explosion, those in the experimental condition continued to take as much risk as before a balloon explosion. Thus, receiving direct encouragement from a risk-promoting robot seemed to override participants' di-

rect experiences and feedback. Reinforcement learning models (Biele et al., 2009) have described the influence of others' recommendations on decision-making with outcome-bonus models. In these models, rewards from a choice that was recommended by others produce more positive reinforcements than rewards from nonrecommended options. Intriguingly, and in line with the findings of the current study, negative experiences with a recommended option inhibit the choice of this option less than negative experiences with nonrecommended options. However, humans are biased in whom they trust for advice, preferring, for example, reliable or prestigious advisers (Rendell et al., 2011). Indeed, our results indicate that participants in the experimental condition had an overall positive impression of the robot adviser and felt safe in its presence, particularly toward the end of the experimental session.

Several limitations should be acknowledged. First, our sample was composed of mostly undergraduate female students. While many other studies relied on university students, previous work has shown that males exhibit higher risk-taking behavior. Thus, it is feasible that our results are conservative by nature and a sample that includes more males would have shown an even greater impact of the robot. Likewise, earlier studies (Gheorghiu et al., 2015; Pradhan et al., 2014; Reynolds et al., 2014; Shepherd et al., 2011; Simons-Morton et al., 2005; Steinberg & Monahan, 2007; Toxopeus et al., 2011) have focused on the impact of peers on adolescent risk taking, as this age group not only tends to be a high-risk taker but more likely to be influenced by peers. Furthermore, we have focused on one type of risk, namely, financial. Whether robots would be able to influence people's risk-taking in other domains – such as ethical, social, or recreational – is an open, and pressing, question. Second, in this study, we only studied the interaction between humans and robots and cannot conclude whether similar results would emerge from human interaction with other artificial intelligence (AI) systems, such as digital assistants or on-screen avatars. With the wide spread of AI technology and its interactions with humans, this is an area that needs urgent attention from the research community.

Finally, here we focused on whether robots can increase risk-taking behavior. We are unable to tell whether they can also lead to reductions in risky behavior (see Supplementary Data[6] for further limitations).

Despite the growing body of research on HRI and its utilization across domains, there is a clear paucity of research examining whether robots can influence human risk-taking behavior by encouraging risky choices. Here, we took the first step in addressing this question. Our da-

[6] https://www.liebertpub.com/doi/suppl/10.1089/cyber.2020.0148/suppl_file/Supp_Data.docx.

ta reveal that HRI could lead to increased risk-taking behavior. On the one hand, our results might raise alarms about the prospect of robots (and other AI agents) causing harm by increasing risky behavior. On the other hand, our data point to the possibility of utilizing robots (and other AI agents) in preventive programs (such as antismoking campaigns in schools), and with hard-to-reach populations, such as addicts.

References

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, Leadership, and Men* (pp. 177-190). Carnegie Press.

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, *1*(1), 71-81.

Biele, G., Rieskamp, J., & Gonzalez, R. (2009). Computational models for the combination of advice and individual learning. *Cognitive science*, *33*(2), 206-242.

Brandstetter, J., Rácz, P., Beckner, C., Sandoval, E. B., Hay, J., & Bartneck, C. (2014). A peer pressure experiment: Recreation of the Asch conformity experiment with robots. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 1335-1340). IEEE.

Chou, E. Y., & Nordgren, L. F. (2017). Safety in numbers: Why the mere physical presence of others affects risk-taking behaviors. *Journal of Behavioral Decision Making*, *30*(3), 671-682.

Cottrell, N. B., Wack, D. L., Sekerak, G. J., & Rittle, R. H. (1968). Social facilitation of dominant responses by the presence of an audience and the mere presence of others. *Journal of personality and social psychology*, *9*(3), 245.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the european economic association*, *9*(3), 522-550.

Gardner, M., & Steinberg, L. (2005). Peer influence on risk taking, risk preference, and risky decision making in adolescence and adulthood: an experimental study. *Developmental psychology, 41*(4), 625.

Gheorghiu, A., Delhomme, P., & Felonneau, M. L. (2015). Peer pressure and risk taking in young drivers' speeding behavior. *Transportation research part F: traffic psychology and behaviour*, *35*, 101-111.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*(5812), 619-619.

Guerin, B., & Innes, J. M. (1984). Explanations of social facilitation: A review. *Current psychological research & reviews*, *3*(2), 32-52.

Hopko, D. R., Lejuez, C. W., Daughters, S. B., Aklin, W. M., Osborne, A., Simmons,

B. L., & Strong, D. R. (2006). Construct validity of the balloon analogue risk task (BART): Relationship with MDMA use by inner-city drug users in residential treatment. *Journal of Psychopathology and Behavioral Assessment, 28*(2), 95-101.

Lejuez, C. W., Aklin, W. M., Jones, H. A., Richards, J. B., Strong, D. R., Kahler, C. W., & Read, J. P. (2003). The balloon analogue risk task (BART) differentiates smokers and nonsmokers. *Experimental and Clinical Psychopharmacology, 11*(1), 26.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., Strong, D. R., & Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied, 8*(2), 75.

Lejuez, C. W., Simmons, B. L., Aklin, W. M., Daughters, S. B., & Dvir, S. (2004). Risk-taking propensity and risky sexual behavior of individuals in residential substance use treatment. *Addictive behaviors, 29*(8), 1643-1647.

Pradhan, A. K., Li, K., Bingham, C. R., Simons-Morton, B. G., Ouimet, M. C., & Shope, J. T. (2014). Peer passenger influences on male adolescent drivers' visual scanning behavior during simulated driving. *Journal of Adolescent Health, 54*(5), S42-S49.

Rendell, L., Fogarty, L., Hoppitt, W. J., Morgan, T. J., Webster, M. M., & Laland, K. N. (2011). Cognitive culture: Theoretical and empirical insights into social learning strategies. *Trends in cognitive sciences, 15*(2), 68-76.

Reynolds, E. K., MacPherson, L., Schwartz, S., Fox, N. A., & Lejuez, C. W. (2014). Analogue study of peer influence on risk-taking behavior in older adolescents. *Prevention Science, 15*(6), 842-849.

Salomons, N., Van Der Linden, M., Strohkorb Sebo, S., & Scassellati, B. (2018). Humans conform to robots: Disambiguating trust, truth, and conformity. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction* (pp. 187-195).

Shepherd, J. L., Lane, D. J., Tapscott, R. L., & Gentile, D. A. (2011). Susceptible to social influence: Risky "driving" in response to peer pressure 1. *Journal of Applied Social Psychology, 41*(4), 773-797.

Simons-Morton, B., Lerner, N., & Singer, J. (2005). The observed effects of teenage passengers on the risky driving behavior of teenage drivers. *Accident Analysis & Prevention, 37*(6), 973-982.

Steinberg, L., & Monahan, K. C. (2007). Age differences in resistance to peer influence. *Developmental psychology, 43*(6), 1531.

Terada, K., Shamoto, T., Ito, A., & Mei, H. (2007). Reactive movements of non-humanoid robots cause intention attribution in humans. In *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 3715-3720). IEEE.

Toxopeus, R., Ramkhalawansingh, R., & Trick, R.L.M. (2011) The influence of passenger-driver interaction on young drivers. *Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. 66-72).

Vollmer, A. L., Read, R., Trippas, D., & Belpaeme, T. (2018). Children conform,

adults resist: A robot group induced peer pressure on normative social conformity. *Science robotics, 3*(21), eaat7111.

World Health Organization. (2018). *Global Status Report on Road Safety 2018*. World Health Organization. https://www.who.int/violence_injury_prevention/road_safety_status/2018/en.

Xu, K., & Lombard, M. (2017). Persuasive computing: Feeling peer pressure from multiple computer agents. *Computers in Human Behavior, 74*, 152-162.

6. A Robot Is Not Worth Another

Exploring Children's Mental State Attribution to Different Humanoid Robots

F. Manzi, G. Peretti, C. Di Dio, A. Cangelosi, S. Itakura, T. Kanda, H. Ishiguro, D. Massaro, A. Marchetti

ABSTRACT

Recent technological developments in robotics have driven the design and production of different humanoid robots. Several studies have highlighted that the presence of human-like physical features could lead both adults and children to anthropomorphize the robots. In the present study we aimed to compare the attribution of mental states to two humanoid robots, NAO and Robovie, which differed in the degree of anthropomorphism. Children aged 5, 7, and 9 years were required to attribute mental states to the NAO robot, which presents more human-like characteristics compared to the Robovie robot, whose physical features look more mechanical. The results on mental state attribution as a function of children's age and robot type showed that 5-year-olds have a greater tendency to anthropomorphize robots than older children, regardless of the type of robot. Moreover, the findings revealed that, although children aged 7 and 9 years attributed a certain degree of human-like mental features to both robots, they attributed greater mental states to NAO than Robovie compared to younger children. These results generally show that children tend to anthropomorphize humanoid robots that also present some mechanical characteristics, such as Robovie. Nevertheless, age-related differences showed that they should be endowed with physical characteristics closely resembling human ones to increase older children's perception of human likeness. These findings have important

---

This chapter was originally published as Manzi, F., Peretti, G., Di Dio, C., Cangelosi, A., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., & Marchetti, A. (2020). A robot is not worth another: Exploring children's mental state attribution to different humanoid robots. *Frontiers in Psychology, 11*, 2011. The datasets generated for this study are available on request to the corresponding author. The studies involving human participants were reviewed and approved by the Ethic Committee, Università Cattolica del Sacro Cuore, Milan, Italy. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. All the authors conceived and designed the experiment. FM and GP conducted the experiments in schools. AM, FM, and GP secured ethical approval. FM and CDD carried out the statistical analyses. All authors contributed to the writing of the manuscript. This publication was granted by Università Cattolica del Sacro Cuore of Milan. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

implications for the design of robots, which also needs to consider the user's target age, as well as for the generalizability issue of research findings that are commonly associated with the use of specific types of robots.

Introduction

Currently, we are witnessing an increasing deployment of social robots (Bartneck & Forlizzi, 2004) in various contexts, from occupational to clinical to educational (Belpaeme et al., 2018; Marchetti et al., in press; Murashov et al., 2016;). Humanoid social robots (HSRs), in particular, have proven to be effective social partners, possibly due to their physical human likeness (Dario et al., 2001). Humanoid social robots can vary in the degree of their anthropomorphic physical characteristics, often depending on the target user (children, adults, elderly, students, clinical populations, etc.) and the context (household, education, commercial, and rehabilitation). For example, the humanoid KASPAR robot that resembles a young child (with face, arms and hands, legs and feet), was specifically built for children with autism spectrum disorder (Dautenhahn et al., 2009; Wainer et al., 2014). In other instances, however, the same HSRs are used both for different purposes and different populations, like the NAO robot, which is largely used both with clinical and non-clinical populations (Begum et al., 2016; Belpaeme et al., 2018; Mubin et al., 2013; Shamsuddin et al., 2012), or the Robovie robot, that is employed both with adults and children (Kahn et al., 2012; Shiomi et al., 2006). A recent review of the literature by Marchetti et al. (2018) showed that different physical characteristics of HSRs may significantly affect the quality of interaction between humans and robots at different ages. The construction of robots that integrate and expand the specific biological abilities of our species led to two different directions in robotic development based on different, though related, theoretical perspectives: developmental cybernetics (DC; Di Dio et al., 2019; Itakura, 2008; Itakura et al., 2008; Kannegiesser et al., 2015; Manzi et al., 2020a; Moriguchi et al., 2011; Okanda et al., 2018; Wang et al., 2020) and developmental robotics (DR; Cangelosi & Schlesinger, 2015, 2018; De La Cruz et al., 2014; Di Dio et al., 2020a,b; Lyon et al., 2016; Morse & Cangelosi, 2017; Vinanzi et al., 2019; Zhong et al., 2019). The first perspective (DC) consists of creating a human-like system, by simulating human psychological processes and prosthetic functions in the robot (enhancing the function and lifestyle of persons) to observe people's behavioral response toward the robot. The second perspective (DR) is related to the development of cognitive neural networks in the robot that would allow it to autonomously gain sensorimotor and mental capabili-

ties with growing complexity, starting from intricate evolutionary principles. From these premises, the next two paragraphs briefly outline current findings concerning the effect that physical features of the HSRs have on human perception, thus outlining the phenomenon of anthropomorphism, and a recent methodology devised to measure it.

Anthropomorphism

Anthropomorphism is a widely observed phenomenon in human-robot interaction (HRI; Airenti, 2015; Fink et al., 2012; Złotowski et al., 2015), and it is also greatly considered in the design of robots (Dario et al., 2001; Bartneck et al., 2009; Kiesler et al., 2008; Sharkey & Sharkey, 2011; Zanatto et al., 2016, 2020). In psychological terms, anthropomorphism is the tendency to attribute human characteristics, physical and/or psychological, to non-human agents (Duffy, 2003; Epley et al., 2007). Several studies have shown that humans may perceive non-anthropomorphic robots as anthropomorphic, such as Roomba (a vacuum cleaner with a semi-autonomous system; Fink et al., 2012). Although anthropomorphism seems to be a widespread phenomenon, the attribution of human traits to anthropomorphic robots is significantly greater compared to non-anthropomorphic robots. A study by Krach et al. (2008) compared four different agents (computer, functional robot, anthropomorphic robot, and human confederate) in a Prisoner's Dilemma Game, and showed that the more the interactive partner displayed human-like characteristics, the more the participants appreciated the interaction and ascribed intelligence to the game partner. What characteristics of anthropomorphic robots (i.e., the HSRs) increase the perception of anthropomorphism? The HSRs can elicit the perception of anthropomorphism mainly at two levels: physical and behavioral (Marchetti et al., 2018). Working on the physical level is clearly easier than on intrinsic psychological features, and – although anthropomorphic physical features of robots are not the only answer to enhance the quality of interactions with humans – the implementation of these characteristics can positively affect HRIs (Duffy, 2003; for a review see Marchetti et al., 2018). It should be stated, however, that extreme human-likeness can result in the known uncanny valley effect, according to which HRIs are negatively influenced by robots that are too similar to the human (MacDorman & Ishiguro, 2006; Mori, 1970; Mori et al., 2012). Thus, the HSRs' appearance represents an important social affordance for HRIs, as further demonstrated by the psychological research on racial and disability prejudice (Macdonald et al., 2017; Manzi et al., 2020b; Sarti et al., 2019; Todd et al., 2011). The anthropomorphic features of the HSRs

can increase humans' perception of humanness, such as mind attribution and personality, and influence other psychological mechanisms and processes (Bartneck et al., 2008; Broadbent et al., 2013; Kiesler & Goetz, 2002; MacDorman et al., 2005; Marchetti et al., 2020; Powers & Kiesler, 2006; Złotowski et al., 2015).

The study of the design of physical characteristics of the HSRs and their classification has been already investigated in HRI, but not systematically. A pioneering study by DiSalvo et al. (2002) explored the perception of humanness using 48 images of different heads of HSRs, and showed that three features are particularly important for the robot's design: the nose, eyes, and mouth. Furthermore, a study by Duffy (2003) categorized different robots' head in a diagram composed of three extremities: 'human head' (as-close-as-possible to a human head), 'iconic head' (a very minimum set of features) and 'abstract head' (a more mechanistic design with minimal human-like aesthetics). Also, in this instance, human likeness was associated with greater mental abilities. Furthermore, a study by MacDorman (2006) analyzed the categorization of 14 types of robots (mainly androids and humanoids) in adults. It was shown that humanoid robots displaying some mechanical characteristics – such as the Robovie robot – were classified average on a 'humanness' scale and rated lower on the uncanny valley scale. Recent studies compared one of the most widely used HSRs, the NAO robot, with different types of robots. It was shown that the NAO robot is perceived less human-like than an android – which is a highly anthropomorphic robot in both appearance and behavior (Broadbent, 2017) –, but more anthropomorphic than a mechanical robot, i.e., the Baxter robot (Yogeeswaran et al., 2016; Zanatto et al., 2019). However, there were no differences in perceived ability to perform physical and mental tasks between NAO and the android (Yogeeswaran et al., 2016), indicating that human-likeness (and not 'human-exactness') is sufficient to trigger the attribution of psychological features to a robot. In addition, a database has recently been created that collects more than 200 HSRs classified according to their level of human likeness (Phillips et al., 2018). In this study the NAO robot was classified with a score of about 45/100, in particular thanks to the characteristics of its face and body. Robovie and other similar robots were classified with a score ranging between 27 and 31/100, deriving mainly from body characteristics. These findings corroborate the hypothesis that NAO and Robovie are two HSRs with different levels of human-likeness due to their physical anthropomorphic features.

The interest in observing the effect of different physical characteristics of robots in terms of attribution of intentions, understanding, and emotions has also been investigated in children (Bumby & Dautenhahn,

1999; Woods et al., 2004; Woods, 2006). In particular, a study by Woods (2006), comparing 40 different robots, revealed that children experience greater discomfort with robots that look too similar to humans, favoring robots with mixed human-mechanical characteristics. These results were confirmed in a recent study by Tung (2016) showing that children preferred robots with not too many human-like features over robots with many human characteristics. Overall, these results suggested that an anthropomorphic design of HSRs may increase children's preference toward them. Still, an excessive implementation of human features can negatively affect the attribution of positive qualities to the robot, again in line with the uncanny valley effect above.

Attribution of mental states

Different scales were developed to measure psychological anthropomorphism toward robots in adults. These scales typically assess attribution of intelligence, personality and emotions, only to mention a few. In particular, the attribution of internal states to the robot, i.e., to have a mind, is widely used and very promising in HRI (Broadbent et al., 2013; Stafford et al., 2014).

In psychology, the ability to ascribe mental states to others is defined as the Theory of Mind (ToM). ToM is the ability to understand one's own and others' mental states (intentions, emotions, desires, beliefs), and to predict and interpret one's own and others' behaviors on the basis of such meta-representation (Perner & Wimmer, 1985; Premack & Woodruff, 1978; Wimmer & Perner, 1983). ToM abilities develop around four years of age, becoming more sophisticated with development (Wellman et al., 2001). ToM is active not only during humans' relationships but also during interactions with robots (for a review, see Marchetti et al., 2018).

Recent studies have shown that adults tend to ascribe greater mental abilities to robots that have a human appearance (Hackel et al., 2014; Martini et al., 2016). This tendency to attribute human mental states to robots was also observed in children. Generally, children are inclined to anthropomorphize robots by attributing psychological and biological characteristics to them (Katayama et al., 2010; Okanda et al., 2019). Still, they do differentiate between humans and robots' abilities. A pioneering study by Itakura (2008) investigating the attribution of mental verbs to a human and a robot showed that children did not attribute the epistemic verb 'think' to the robot. More recent studies have further shown that already from three years of age, children fairly differentiate a human from a robot in terms of mental abilities (Di Dio et al., 2020a), al-

though younger children appear to be more inclined to anthropomorphize robots compared to older children. This effect may be due to the phenomenon of animism, particularly active at three years of age (Di Dio et al., 2020a,b).

Aim of the study

The present study aimed to investigate the attribution of mental states (AMS) in children aged 5-9 years to two humanoid robots, NAO and Robovie, varying in their anthropomorphic physical features (DiSalvo et al., 2002; Duffy, 2003). Differences in the attribution of mental qualities to the two robots were then explored using the robots' degree of physical anthropomorphism and the child's chronological age. The two humanoid robots, NAO and Robovie, have been selected for two main reasons: (1) in relation to their physical appearance, both robots belong to the category of HSRs, but differ for their degree of anthropomorphism (for a detailed description of the robots, see section 'Materials'); (2) both robots are largely used in experiments with children (Di Dio et al., 2020a,b; Cangelosi & Schlesinger, 2015, 2018; Hood et al., 2015; Kahn et al., 2012; Kanda et al., 2002; Okumura et al., 2013a,b; Kose & Yorganci, 2011; Shamsuddin et al., 2012; Tielman et al., 2014).

In light of previous findings associated with the use of these specific robots described above, we hypothesized the following: (1) independent of age, children would distinguish between humans and robots in terms of mental states by ascribing lower mental attributes to the robots; (2) children would tend to attribute greater mental qualities to NAO compared to Robovie because of its greater human-likeness; and (3) younger children would tend to attribute more human characteristics to robots (i.e., to anthropomorphize more) than older children.

Materials and Methods

Participants

Data were acquired on 189 Italian children from kindergarten and primary school age. The children were divided into three age groups for each robot as follows: (1) for the NAO robot, 5 years (N = 24, 13 females; M = 68.14; SD = 3.67); 7 years (N = 25, 13 females; M = 91.9; SD = 3.43); and 9 years (N = 23, 12 females; M = 116.38, SD = 3.91); (2) for the Robovie robot, 5 years (N = 33, 13 females; M = 70.9, SD = 2.95); 7 years (N = 49, 26 females; M = 93.4, SD = 3.62); and 9 years (N = 35, 15 females; M

= 117.42, SD = 4.44). The initial inhomogeneity between sample sizes in the NAO and Robovie conditions were corrected by the random selection of children in the Robovie condition, caring to balance by gender. Accordingly, the sample for the Robovie condition used for statistical analysis was composed as follows: 5 years (N = 24, 8 females; M = 70.87, SD = 3.1); 7 years (N = 25, 14 females; M = 92.6, SD = 3.73); and 9 years (N = 23, 10 females; M = 117.43, SD = 4.62). The children's parents received a written explanation of the procedure of the study, the measurement items, and gave their written consent. The children were not reported by teachers or parents for learning and/or socio-relational difficulties. The study was approved by the Local Ethic Committee (Università Cattolica del Sacro Cuore, Milan).

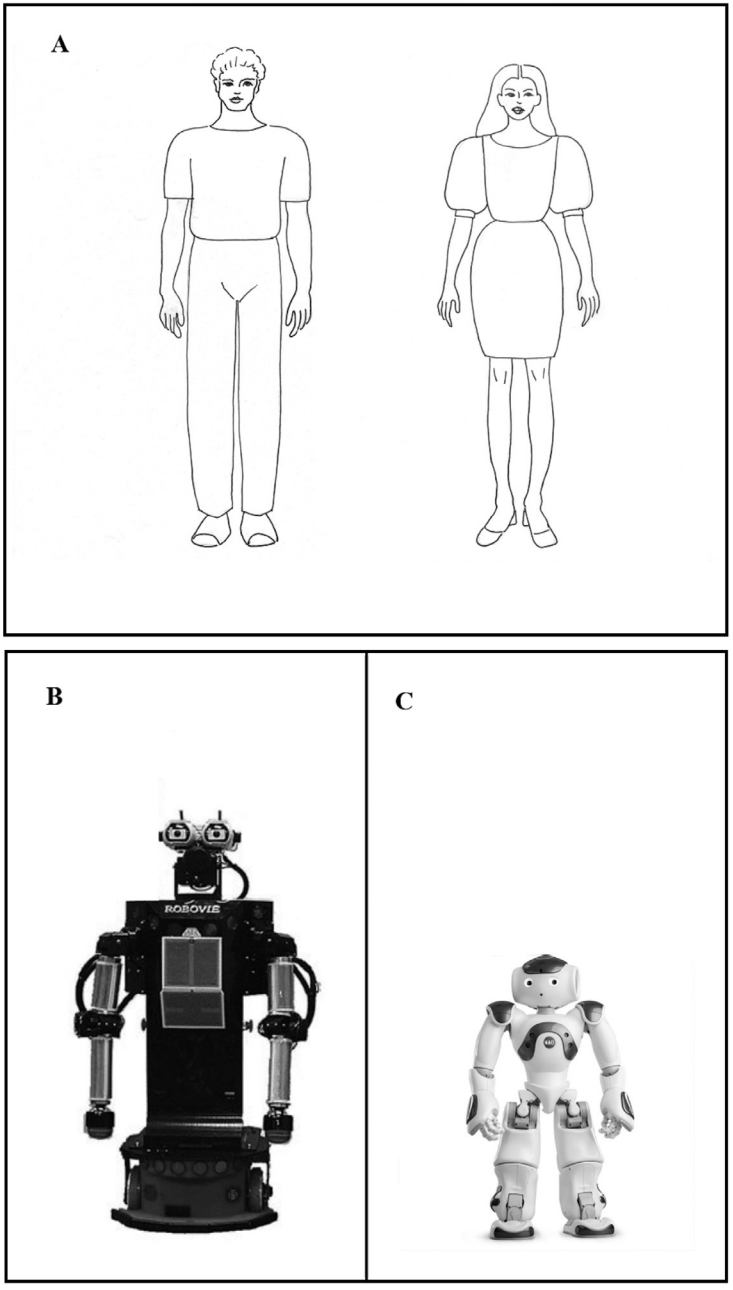
Materials, task, and procedure

Materials. The two HSRs selected for this study were the Robovie robot (Hiroshi Ishiguro Laboratories, ATR; Figure 1B) and the NAO robot (Aldebaran Robotics, Figure 1C). We chose these two robots because, although they both belong to the category of HSRs, they differ in their degree of anthropomorphic features (DiSalvo et al., 2002; Duffy, 2003; MacDorman, 2006; Zhang et al., 2008; Phillips et al., 2018). Robovie is a HSR with more abstract anthropomorphic features: no legs but two driving wheels to move, two arms without hands. In particular, the head can be considered 'abstract' (Duffy, 2003) because of two important human-like features: two eyes and a microphone that looks like a mouth (DiSalvo et al., 2002). Robovie is an HSR that can be rated as average in the continuum of mechanical-humanlike (Ishiguro et al., 2001; Kanda et al., 2002; MacDorman, 2006). NAO is a HSR with more pronounced anthropomorphic features compared to Robovie: two legs, two arms, and two hands with three moving fingers (Figure 1C). Besides, the face can be classified as 'iconic' and consists of three cameras suggesting two eyes and a mouth. However, considering the whole body and the more detailed shape of the face, NAO is a HSR that can be rated as more human-like than Robovie (DiSalvo et al., 2002; MacDorman, 2006; Phillips et al., 2018).

Attribution of Mental States. The AMS questionnaire[1] is a measure of mental states that participants attribute to when they look at images depicting specific characters, in this case a human (female or male based on the participant's gender; Figure 1A), and, according to the group condi-

[1] http://www.teoriadellamente.it, "Strumenti" section.

Figure 1 - *The AMS images: (A) the human condition (male and female), (B) Robovie robot, and (C) the NAO robot*



tion, the Robovie or the NAO robot (Figures 1B and C). The AMS questionnaire was inspired by the methodology described in Martini et al. (2016) and is already used in several experiments with children (Di Dio et al., 2019, 2020a,b; Manzi et al., 2017). The construction of the content of the questionnaire is based on the theoretical model of Slaughter et al. (2009) on the categorization of children's mental verbs resulting from communication exchanges between mother and child. This classification divides mental verbs into four categories: perceptive, volitional, cognitive, and dispositional. For the creation of the AMS questionnaire an additional category related to imaginative verbs has been added. We considered it necessary to distinguish between cognitive, epistemic, and imaginative states, since – especially for the robot – this specification enables the analysis of different psychological processes in terms of development. The AMS therefore consists of five dimensions: Perceptive, Emotive, Desires and Intentional, Imaginative, and Epistemic.

The human condition was used as a baseline measure to evaluate children's ability to attribute mental states. In fact, as described in the results below, children scored quite high when ascribing mental attributes to the human character, thus supporting children's competence in performing the mental states attribution task. Also, the human condition was used as a comparison measure against which the level of psychological anthropomorphism of NAO and Robovie was evaluated. The Cronbach's alfa for each category is as follows: Perceptive ($\alpha = 0.8$), Emotive ($\alpha = 0.8$), Desires and Intentional ($\alpha = 0.8$), Imaginative ($\alpha = 0.8$), and Epistemic ($\alpha = 0.7$).

Children answered 25 questions grouped into the five different state categories described above (see Appendix 1 for the specific items). The child had to answer 'Yes' or 'No' to each question, obtaining 1 when the response is 'Yes' and 0 when the response is 'No'. The sum of all responses (range = 0-25) gave the total score ($\alpha = 0.9$); the five partial scores were the sum of the responses within each category (range = 0-5).

Procedure. The children were tested individually in a quiet room inside their school. Data acquisition was carried out by a single researcher during the normal school activities.

The experimenter showed each child the image on a paper depicting a human – gender matched – and one of the two robots, NAO or Robovie. The presentation order of the image – human and robot – was randomized. Afterward, the experimenter asked children the questions on the five categories of the AMS (Perceptive, Emotive, Intentions and Desires, Imaginative, and Epistemic). The presentation order of the five categories was also randomized. The total time required to complete the test was approximately 10 min.

Results

Data analysis

To evaluate the effect of age, gender, states, agent, and type of robots on children's mental state attribution to robots, a GLM analysis was carried out with five levels of *states* (Perceptive, Emotive, Intentions and Desires, Imaginative, and Epistemic) and two levels of *agent* (Human, Robot) as within-subjects factors, and *age* (5-, 7-, 9-year-olds), *gender* (Male, Female) and *robot* (Robovie, NAO) as the between-subjects factor. The Greenhouse-Geisser correction was used for violations of Mauchly's Test of Sphericity ($p < 0.05$). *Post hoc* comparisons were Bonferroni corrected.

Results

The results showed (1) a main effect of *agent*, $F(1, 126) = 570.9$, $p < 0.001$, *partial*-$\eta^2 = 0.819$, $\delta = 1$, indicating that children attributed greater mental states to the human ($M = 4.6$, SD = 0.27) compared to the robot ($M = 2.7$, SD = 0.21; *M*diff = 1.75, SE = 0.087); (2) a main effect of *states*, $F(4,504) = 40.33$, $p < 0.001$, *partial*-$\eta^2 = 0.243$, $\delta = 1$, mainly indicating that children attributed greater intention and desires and lower imaginative states (for a full description of the statistics, see Table 1); (3) a main effect of *robot*, $F(1,126) = 39.4$, $p < 0.001$, *partial*-$\eta^2 = 0.238$, $\delta = 1$, showing that children attributed greater mental states to NAO ($M = 3.98$, SD = 0.17) compared to Robovie ($M = 3.4$, SD = 0.14; *M*diff = 0.568, SE = 0.099).

A two-way interaction was also found between (1) *states* and *agent*, $F(1,126) = 16.51$, $p < 0.001$, *partial*-$\eta^2 = 0.183$, $\delta = 1$ (for a detailed description of the differences see Table 2 on page 168), and (2) *agent* and *age*, $F(2,126) = 25.17$, $p < 0.001$, *partial*-$\eta^2 = 0.285$, $\delta = 1$, showing that 5-year-old children attributed greater mental states to the robotic agents compared to older children (see Table 2).

Additionally, we found a three-way interaction between *states*, *age*, and *robot*, $F(8,126) = 4.95$, $p < 0.001$, *partial*-$\eta^2 = 0.073$, $\delta = 1$. The planned comparisons on the three-way interaction revealed that children attributed greater mental states to NAO compared to Robovie, with the youngest children differentiating on the Perceptive and Epistemic dimensions, and with this difference spreading to all dimensions (but imaginative) in the older children (see Figure 2 on page 170).

Table 1 - *Statistics comparing the attribution of all AMS dimensions (Perceptive, Emotive, Intentions and Desires, Imaginative, Epistemic)*

Dimension	*Mental States*	*Mdiff*	*Err. Stan.*	*Sign.*
Perceptive	Emotive	.203	.08	.122
	Int&Des	**-.354***	.071	**.000**
	Imaginative	**.525***	.075	**.000**
	Epistemic	-0,089	.069	1
Emotive	Perceptive	-0,203	.08	.122
	Int&Des	**-.557***	.074	**.000**
	Imaginative	**.322***	.072	**.000**
	Epistemic	**-.292***	.079	**.004**
Int&Des	Perceptive	**.354***	.071	**.000**
	Emotive	**.557***	.074	**.000**
	Imaginative	**.879***	.071	**.000**
	Epistemic	**.265***	.067	**.001**
Imaginative	Perceptive	**-.525***	.065	**.000**
	Emotive	**-.322***	.070	**.000**
	Int&Des	**-.879***	.068	**.000**
	Epistemic	**-.614***	.074	**.000**
Epistemic	Perceptive	.089	.065	**1**
	Emotive	**.292***	.070	**.004**
	Int&Des	**-.265***	.068	**.001**
	Imaginative	**.614***	.074	**.004**

Based on estimated marginal averages
* The average difference is significant at the level of 0.05.

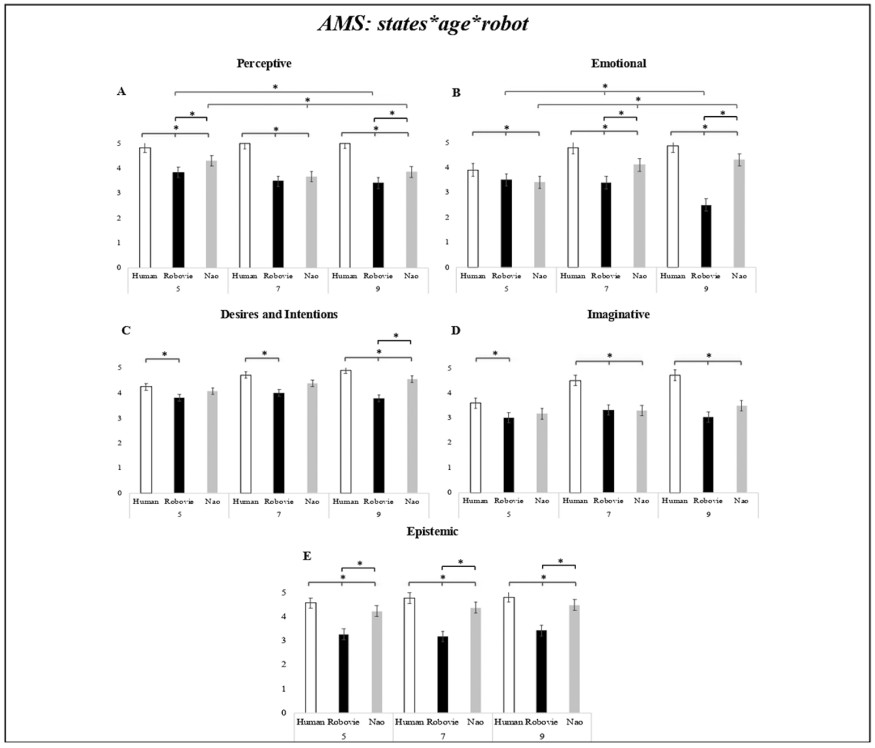
Table 2 - *Statistics comparing the attribution of all AMS dimensions (Perceptive, Emotive, Intentions and Desires, Imaginative, Epistemic) and the AMS for the two agents (Human, Robot) across ages (5-, 7- and 9-years)*

		Human			*Robot*		
	Age	*Mdiff*	*Err. Stan.*	*Sign.*	*Mdiff*	*Err. Stan.*	*Sign.*
	5 vs 7	**-.558***	.103	**.000**	**.443***	.182	**.05**
	5 vs 9	**-.558***	.104	**.000**	**.620***	.183	.003
	7 vs 9	-.108	.101	.866	.177	.179	.97
State	*Dimensions*	*Mdiff*	*Err. Stan.*	*Sign.*	*Mdiff*	*Err. Stan.*	*Sign.*
Perceptive	Emotive	**.405***	.202	**.608**	7,63E-05	-.351	.351
	Int&Des	**.307***	.121	**.493**	**-.015***	-1.35	**-.681**
	Imaginative	**.674***	.458	**.89**	**.376***	.048	**.704**
	Epistemic	**.218***	.087	**.35**	**-.396***	-.758	**-.035**
Emotive	Perceptive	**-.405***	-.608	**-.202**	-7,63E-05	-.351	.351
	Int&Des	-.098	-.323	.126	**-1.016***	-1.366	**-.665**
	Imaginative	**.269***	.046	**.492**	.376*	.052	**.7**
	Epistemic	-.187	-.401	.028	**-.396***	-.767	**-.025**

		Human			*Robot*		
	Age	*Mdiff*	*Err. Stan.*	*Sign.*	*Mdiff*	*Err. Stan.*	*Sign.*
Int&Des	Perceptive	**-.307***	-.493	**-.121**	**1.015***	.681	**1.35**
	Emotive	.098	-.126	.323	**1.016***	.665	**1.366**
	Imaginative	**.367***	.119	**.616**	**1.391***	1.078	**1.705**
	Epistemic	-.088	-.272	.096	**.619***	.296	**.942**
Imaginative	Perceptive	**-.674***	-.89	**-.458**	**-376***	-.704	**-.048**
	Emotive	**-.269***	-.492	-.046	**-.376***	-.7	**-.052**
	Int&Des	**-.367***	-.616	**-.119**	**-1.391***	-1.705	**-1.078**
	Epistemic	**-.456***	-.681	**-.23**	**-.772***	-1.112	**-.432**
Epistemic	Perceptive	**-.218***	-.35	**-.087**	**.396***	.035	**.758**
	Emotive	.187	-.028	.401	**.396***	.025	.767
	Int&Des	.088	-.096	.272	**-.619***	-.942	**-.296**
	Imaginative	**.456***	.23	**.681**	**.772***	.432	**1.112**

Based on estimated marginal averages
* The average difference is significant at the level of 0.05.

Figure 2 (A-E) - *Children's scores on the attribution of mental states (AMS) scale*



Note: AMS mean scores for the Human (white bar), for Robovie robot (black bar), and NAO robot (gray bar) for each state (Perceptive, Emotions, Intentions and Desires, Imagination, and Epistemic) as a function of age group (5-, 7-, and 9-year-olds). The bars represent the standard error of the mean. *Indicates significant differences.

Discussion and Conclusion

Discussion

In the present study we compared the AMS in children aged 5-9 years between two HSRs, NAO and Robovie, also with respect to a human. The aim was to explore children's patterns of mental attribution to different types of HSRs, varying in their degree of physical anthropomorphism, from a developmental perspective.

Our results on the AMS to the human and robot generally confirmed

the tendency of children to ascribe lower human mental qualities to the robots, thus supporting previous findings (Manzi et al., 2017; Di Dio et al., 2018, 2019, 2020a,b). In addition, children generally attributed greater mental states to the NAO robot than to the Robovie robot, although differences were found in the quality of mental states attribution as a function of age, with older children discriminating more between the types of robots that the younger ones. As a matter of fact, the important role played by the type of robot in influencing children's AMS can be appreciated by evaluating differences in state attribution developmentally.

Firstly, 5-year-old children generally attributed greater human-like mental states to the robotic agents compared to older children. Additionally, while 5-year-old children discriminated between robots' mental attribution only on the perceptive and epistemic dimensions – with the NAO robot being regarded as more anthropomorphic than Robovie –, children aged 7 and 9 years were particularly sensitive to the type of robots, and attributed greater mental states to NAO than Robovie on most of the tested mental state dimensions. From a developmental perspective, the tendency of younger children to anthropomorphize HSRs could be reasonably explained by the phenomenon of animism (Piaget, 1929). Already Piaget in 1929 suggested that children younger than 6 years tend to attribute a consciousness to objects, i.e., the phenomenon of animism, and that this fades around 9 years of age. Recently, this phenomenon has been defined as a cognitive error in children (Okanda et al., 2019), i.e., animism error, characterized by a lack of differentiation between living and non-living things. In this respect, several studies showed that, although children are generally able to discriminate between humans and robots, children aged 5-6 years tend to overuse animistic interpretations for inanimate things, and to attribute biological and psychological properties to robots (Katayama et al., 2010; Di Dio et al., 2019, 2020a, b), in line with the results of this study. Interestingly, we further found a difference in emotional attribution to NAO between 5-year-olds and 7- and 9-year-old children: younger children attributed lower emotions to NAO compared to the older ones. This result may seem counterintuitive in light of what we discussed above; however, by finely looking at the scores obtained from the 5-year-olds for each single emotional question, we found that younger children attributed significantly lower negative emotions to NAO compared to the other age groups, favoring positive emotions ($\chi^2 < 0.01$). This resulted in an overall decrease of scores in the emotional dimension for the young children. Therefore, not only does this result not contradict the idea of a greater tendency to anthropomorphize robots in younger children compared to older ones, but also highlights that 5-years-olds perceive NAO as a posi-

tive entity that cannot express negative emotions such as anger, sadness, and fear: the 'good' play-partner.

From the age of 7, children's belief of the robots' mind is significantly affected by a sensitivity to the type of the robot, as shown by differences between NAO and Robovie on most mental dimensions, except for Imaginative. The lack of differences between robots on the Imaginative dimension (for all age-groups), which encompasses psychological processes like pretending, and making jokes, appears to be regarded by children as a human prerogative. Interestingly, this result supports findings from a previous study (Di Dio et al., 2018) that compared 6-year-old children's mental state attribution to different entities (human, dog, robot, and God). Also, in that study, imagination was specific to the human entity.

Generally, the findings for older children indicate that the robot's appearance does affect mental state attribution to the robot, and this is increasingly evident with age. However, the judgment of older children could also be significantly influenced by the robot's behavioral characteristics, as demonstrated in a long-term study conducted with children aged 10-12 years (Ahmad et al., 2016). In this study, children played a snakes and ladders game with a NAO robot three times across 10 days, whose behavior in terms of personality for a social robot in education was adapted to maintain and create long-term engagement and acceptance. It was found that children positively reacted to the use of the robot in education, stressing a need to implement robots that are able to adapt based on previous experiences in real time. Of course, this is very much in line with the great vision of disciplines such DR (Cangelosi & Schlesinger, 2015) and DC (Itakura, 2008). In this respect, it is also important to consider further aspects related to the effectiveness in human relations of constructs such as understanding the perspective of others (e.g., Marchetti et al., 2018) and empathy, on which several research groups are actively working. For example, in an exploratory study Serholt et al. (2014) highlighted the perceived need both for teachers and learners to deal with robots showing such a competence.

In the same vein, other studies that used Robovie as an interactive partner in educational contexts, have also shown that when the robot is programed to facilitate interactional dynamics with children, it can be considered by the children as a group member and even part of the friendship circle. In these studies, the robot is typically programed to act as an effective social communicative partner using strategies, like calling children by their name, or adapting the interactive behaviors for each child by means of behavioral patterns drawn from developmental psychology (Kanda et al., 2007; see also, Kahn et al., 2012). The study by Kahn et al. (2012) further showed that after interacting with Robovie,

most children believed that Robovie had mental states (e.g., was intelligent and had feelings) and was a social being (e.g., could be a friend, offer comfort, and be trusted with secrets).

The above studies highlight the prospective use of robots, particularly in the educational field. However, in reality, today's robots are not yet able to sustain autonomous behavior in the long term, even though research is actively laying a good foundation for this. What we can certainly work on with direct effects on children's perception of the mental abilities of robots are their physical attributes. By outlining differences in mental states attribution to different types of humanoid robots across ages based on robots' physical appearance, our findings could help map the design of humanoid robots for children: in early ages, robots can display more abstract and mechanical features (possibly also due to the phenomenon of animism as described above); conversely, in older ages, the tendency to anthropomorphize robots is at least partially affected by the design of the robot. However, it has to be kept in mind that excessive human-likeness may be felt as uncomfortable, as suggested by findings showing that children experience less discomfort with robots displaying both human and mechanical features compared to robots whose physical features markedly evoke human ones (Bumby & Dautenhahn, 1999; Woods et al., 2004; Woods, 2006). Excessive resemblance to the human triggers the uncanny valley effect (the more the appearance of robots is similar to humans, the higher the sense of eeriness). These data suggest that a well-designed HSR for children should combine both human and mechanical dimensions, which, in our study, seems to be better represented by the NAO robot.

Conclusion

This study enabled us to analyze the AMS to two types of HSRs, highlighting how different types of robots can evoke different attributions of mental states in children. More specifically, our findings suggest that children's age is an important factor to consider when designing a robot, and provided us with at least two important insights associated with the phenomenon of anthropomorphism from a development perspective, and the design of HSRs for children. Anthropomorphism seems to be a widespread phenomenon in 5-year-olds, while it becomes more dependent on physical features of the robot in older children, with a preference ascribed to the NAO robot that is perceived as more human-like. This effect may then influence the design of robots, which can be more flexible in terms of physical features, as with Robovie, when targeted to young children.

Overall, our results suggest that the assessment of HSRs in terms of mental states attribution may represent a useful measure for studying the effect of different robots' design for children. However, it has to be noted that the current results involved only two types of HSRs. Therefore, future studies will have to evaluate the mental attribution to a greater variety of robots by also comparing anthropomorphic and non-anthropomorphic robots, and across different cultures. In addition, in future studies it will be important to assess children's socio-cognitive abilities such as language, executive functions, and ToM, to analyze the effect of these abilities on the AMS to robots developmentally. Finally, this study explored the mental attributions through images depicting robots. Future studies should include a condition where children interact with the robots *in vivo* to explore the intersectional effect between the robot's physical appearance and its behavioral patterns. This would enable us to highlight the relative weight of each factor on children's perception of the robots' mental competences.

APPENDIX 1

Attribution of Mental States (AMS)

I will show you an image of a girl/boy/robot (to be selected depending on condition). I will ask you some questions about her/him/it (depending on condition). You can answer Yes or No to the questions.

Dimensions (5) and Questions (25)

Perceptive
Do you think she/he/it can smell?
Do you think she/he/it can see?
Do you think she/he/it can taste?
Do you think she/he/it can hear?
Do you think she/he/it can feel hot or cold?
Emotive
Do you think she/he/it can get angry?
Do you think she/he/it can be scared?
Do you think she/he/it can be happy?
Do you think she/he/it can be surprised?
Do you think she/he/it can be sad?

Intentions and Desires
Do you think she/he/it may have the intention to do something?
Do you think she/he/it might want to do something?
Do you think she/he/it might be willing to do something?
Do you think she/he/it can make a wish?
Do you think she/he/it might prefer one thing over another?
Imaginative
Do you think she/he/it can tell a lie?
Do you think she/he/it can pretend?
Do you think she/he/it can imagine?
Do you think she/he/it can make a joke?
Do you think she/he/it can dream?
Epistemic
Do you think she/he/it can understand?
Do you think she/he/it can make a decision?
Do you think she/he/it can learn?
Do you think she/he/it can teach?
Do you think she/he/it can think?

References

Ahmad, M. I., Mubin, O., & Orlando, J. (2016, November). Children views' on social robot's adaptations in education. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction* (pp. 145-149).

Bartneck, C., Croft, E., & Kulic, D. (2008). Measuring the anthropomorphism, animacy, likeability, perceived intelligence and perceived safety of robots. *Proceedings of the Metrics for Human-Robot Interaction Workshop in affiliation with the 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI 2008), Technical Report 471*, Amsterdam (pp. 37-44).

Bartneck, C., & Forlizzi, J. (2004). A design-centered framework for social human-robot interaction. In *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759)* (pp. 591-594). IEEE (September).

Begum, M., Serna, R. W., & Yanco, H. A. (2016). Are robots ready to deliver autism interventions? A comprehensive review. *International Journal of Social Robotics*, 8(2), 157-181.

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science robotics*, *3*(21), eaat5954.

Beran, T. N., Ramirez-Serrano, A., Kuzyk, R., Fior, M., & Nugent, S. (2011). Understanding how children understand robots: Perceived animism in child-robot interaction. *International Journal of Human-Computer Studies,* 69, 539-550.

Broadbent, E. (2017). Interactions with robots: The truths we reveal about ourselves. *Annual review of psychology,* 68, 627-652.

Broadbent, E., Kumar, V., Li, X., Sollers, J., Stafford, R. Q., MacDonald, B.A., & Wegner, D.M. (2013). 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.

Bumby, K., & Dautenhahn, K. (1999). Investigating children's attitudes towards robots: A case study. In *Proc. CT99, The Third International Cognitive Technology Conference* (pp. 391-410) (August).

Cangelosi, A., & Schlesinger, M. (2015). *Developmental Robotics: From Babies to Robots.* MIT Press.

Cangelosi, A., & Schlesinger, M. (2018). From babies to robots: The contribution of developmental robotics to developmental psychology. *Child Development Perspectives, 12*(3), 183-188.

Dario, P., Guglielmelli, E., & Laschi, C. (2001). Humanoids and personal robots: Design and experiments. *Journal of robotic systems, 18*(12), 673-690.

Dautenhahn, K., Nehaniv, C. L., Walters, M. L., Robins, B., Kose-Bagci, H., Mirza, N. A., & Blow, M. (2009). KASPAR – a minimally expressive humanoid robot for human-robot interaction research. *Applied Bionics and Biomechanics, 6*(3-4), 369-397.

De La Cruz, V. M., Di Nuovo, A., Di Nuovo, S., & Cangelosi, A. (2014). Making fingers and words count in a cognitive robot. *Frontiers in Behavioral Neuroscience, 8,* 13.

Di Dio, C., Isernia, S., Ceolaro, C., Marchetti, A., & Massaro, D. (2018). Growing up thinking of God's beliefs: Theory of Mind and ontological knowledge. *Sage Open,* 1-14.

Di Dio, C., Manzi, F., Itakura, S., Kanda, T., Hishiguro, H., Massaro, D., & Marchetti, A. (2019). It does not matter who you are: Fairness in pre-schoolers interacting with human and robotic partners. *International Journal of Social Robotics,* 1-15.

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., & Marchetti, A. (2020a). Come i bambini pensano alla mente del robot: il ruolo dell'attaccamento e della Teoria della Mente nell'attribuzione di stati mentali ad un agente robotico. *Sistemi Intelligenti,* 41-56.

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., & Marchetti, A. (2020b). Shall I trust you? From child human-robot interaction to trusting relationships. *Frontiers in* Psychology, *11,* 469.

DiSalvo, C., & Gemperle, F. (2003). From seduction to fulfillment: The use of anthropomorphic form in design. In *Proceedings of the 2003 International Conference on Designing Pleasurable Products and Interfaces* (pp. 67-72). ACM, New York (June).

DiSalvo, C.F., Gemperle, F., Forlizzi, J., & Kiesler, S. (2002). All robots are not created equal: The design and perception of humanoid robot heads [Conference Paper].

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems, 42*, 177-190.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological review, 114*(4), 864.

Fink, J., Mubin, O., Kaplan, F., & Dillenbourg, P. (2012). Anthropomorphic language in online forums about Roomba, AIBO and the iPad. In *2012 IEEE Workshop on Advanced Robotics and its Social Impacts* (ARSO) (pp. 54-59). IEEE (May).

Hackel, L.M., Looser, C.E., & Van Bavel, J.J. (2014). Group membership alters the threshold for mind perception: The role of social identity collective identification and intergroup threat. *Journal of Experimental Social Psychology, 52*, 15-23.

Hood, D., Lemaignan, S., & Dillenbourg, P. (2015). When children teach a robot to write: An autonomous teachable humanoid which uses simulated handwriting. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 83-90) (March).

Ishiguro, H., Ono, T., Imai, M., Kanda, T., & Nakatsu, R. (2001). Robovie: An interactive humanoid robot. *International Journal of Industrial Robot, 28*(6), 498-503

Itakura, S. (2008). Development of mentalizing and communication: From viewpoint of developmental cybernetics and developmental cognitive neuroscience. *IEICE Transactions of Communications, 91-B*(7), 2109-2117.

Itakura, S., Ishida, H., Kanda, T., Shimada, Y., Ishiguro, H., & Lee, K. (2008). How to build an intentional android: infants' imitation of a robot's goal-directed actions. *Infancy, 13*(5), 519-532.

Kahn Jr, P. H., Kanda, T., Ishiguro, H., Freier, N. G., Severson, R. L., Gill, B. T., Ruckert, J. H., & Shen, S. (2012). "Robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot. *Developmental psychology, 48*(2), 303.

Kanda, T., Ishiguro, H., Ono, T., Imai, M., & Mase, K. (2002). Development and evaluation of an interactive robot "Robovie". In *IEEE International Conference on Robotics and Automation* (pp. 1848-1855). IEEE.

Kanda, T., Sato, R., Saiwaki, N., & Ishiguro, H. (2007). A two-month field trial in an elementary school for long-term human-robot interaction. *IEEE Transactions on robotics, 23*(5), 962-971.

Kannegiesser, P., Itakura, S., Zhou, Y., Kanda, T., Ishiguro, H., & Hood, B. (2015). The role of social eye-gaze in children's and adult's ownership attributions to robotic agents in three cultures. *Interaction Studies, 16*, 1-28.

Katayama N., Katayama, J. I., Kitazaki, M., & Itakura, S. (2010). Young children's folk knowledge of robots. *Asian Cult History, 2*(2), 111.

Kiesler, S., & Goetz, J. (2002). Mental models of robotic assistants. In *Proceedings of the CHI'02 Extended Abstracts on Human Factors in Computing Systems* (New York, NY: ACM), 576-577.

Kose, H., & Yorganci, R. (2011). Tale of a robot: Humanoid robot assisted sign

language tutoring. In *2011 11th IEEE-RAS International Conference on Humanoid Robots* (pp. 105-111). IEEE (October).

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PloS One*, *3*(7).

Lyon, C. Nehaniv, C. L., Saunders, J., Belpaeme, T., Bisio, A., Fischer, K., … & Cangelosi, A. (2016). Embodied language learning and cognitive bootstrapping: Methods and design principles. *International Journal of Advanced Robotic Systems*, *13*(3), 105.

Macdonald, S. J., Donovan, C., & Clayton, J. (2017). The disability bias: Understanding the context of hate in comparison with other minority populations. *Disability & Society*, *32*(4), 483-499.

MacDorman, K. F. (2006). Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. In *ICCS/CogSci-2006 long Symposium: Toward Social Mechanisms of Android Science* (pp. 26-29).

MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, *7*(3), 297-337.

MacDorman, K. F., Minato, T., Shimada, M., Itakura, S., Cowley, S., & Ishiguro, H. (2005). Assessing human likeness by eye contact in an android testbed. In *Proceedings of the XXVII Annual Meeting of the Cognitive Science Society* (pp. 21-23) (July).

Manzi, F., Massaro, D., Kanda, T., Kanako, T., Itakura, S., & Marchetti, A. (2017). Teoria della Mente, bambini e robot: L'attribuzione di stati mentali. *Proceedings from XXX Congresso Nazionale AIP, Sezione di Psicologia dello Sviluppo e dell'Educazione*, Messina (September 14-16), 65-66.

Marchetti, A., Di Dio, C., Manzi F., & Massaro, D. (2020). Robotics in clinical and developmental psychology. In *Comprehensive Clinical Psychology, 2nd Edition*. Elsevier.

Marchetti, A., Manzi, F., Itakura, S., & Massaro, D. (2018). Theory of Mind and humanoid robots from a lifespan perspective. *Zeitschrift für Psychologie*, *226*, 98-109.

Martini, M. C., Gonzalez, C. A., & Wiese, E. (2016). Seeing minds in others – Can agents with robotic appearance have human-like preferences? *Plos One*, *11*(2), e0146310.

Mori, M. (1970). The uncanny valley. *Energy*, *7*(4), 33-35.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, *19*(2), 98-100.

Moriguchi, Y., Kanda, T., Ishiguro, H., Shimada, Y., & Itakura, S. (2011). Can young children learn words from a robot? *Interaction Studies*, *12*, 107-119.

Morse, A.F., & Cangelosi, A. (2017). Why are there developmental stages in language learning? A developmental robotics model of language development. *Cognitive Science*, *41*, 32-51.

Mubin, O., Stevens, C. J., Shahid, S., Al Mahmud, A., & Dong, J. J. (2013). A review of the applicability of robots in education. *Journal of Technology in Education and Learning*, *1*(209-0015), 13.

Murashov, V., Hearl, F., & Howard, J. (2016). Working safely with robot workers: Recommendations for the new workplace. *Journal of occupational and environmental hygiene*, *13*(3), D61-D71.

Okanda, M., Taniguchi, K., & Itakura, S. (2019). The role of animism tendencies and empathy in adult evaluations of robots. HAI'19 *Proceedings of the 7th International Conference on Human-Agent Interaction* - Kyoto (October 6-10), 51-58.

Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., & Itakura, S. (2013a). Can infants use robot gaze for object learning?: The effect of verbalization. *Interaction Studies*, *14*(3), 351-365.

Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., & Itakura, S. (2013b). Infants understand the referential nature of human gaze but not robot gaze. *Journal for Experimental Child Psychology*, 116, 86-95.

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that..." attribution of second-order beliefs by 5- to 10-year-old children. *Journal for Experimental Child Psychology, 39*(3), 437-471.

Phillips, E., Zhao, X., Ullman, D., & Malle, B. F. (2018, February). 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* (pp. 105-113).

Piaget, J. (1929). *The Child's Conception of the World*. Routledge.

Powers, A., & Kiesler, S. (2006). 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*, (New York, NY: ACM), 218-225.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515-526.

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2012). The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social cognitive and affective neuroscience, 7*(4), 413-422.

Serholt, S., Barendregt, W., Leite, I., Hastie, H., Jones, A., Paiva, A., ... & Castellano, G. (2014, August). Teachers' views on the use of empathic robotic tutors in the classroom. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication* (pp. 955-960). IEEE.

Shamsuddin, S., Yussof, H., Ismail, L. I., Mohamed, S., Hanapiah, F. A., & Zahari, N. I. (2012). Initial response in HRI-a case study on evaluation of child with autism spectrum disorders interacting with a humanoid robot Nao. *Procedia Engineering, 41*, 1448-1455.

Sharkey, A., & Sharkey, N. (2011). Children, the elderly, and interactive robots. *IEEE Robotics & Automation Magazine, 18*(1), 32-38.

Shiomi, M., Kanda, T., Ishiguro, H., & Hagita, N. (2006). Interactive humanoid robots for a science museum. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot interaction* (pp. 305-312) (March).

Stafford, R. Q., MacDonald, B. A., Jayawardena, C., Wegner, D. M., & Broadbent, E. (2014). Does the robot have a mind? Mind perception and attitudes towards robots predict use of an eldercare robot. *International journal of social robotics*, *6*(1), 17-32.

Tielman, M., Neerincx, M., Meyer, J. J., & Looije, R. (2014, March). Adaptive emotional expression in robot-child interaction. In *2014 9th ACM/IEEE International Conference on Human-Robot Interaction* (HRI) (pp. 407-414). IEEE.

Todd, A. R., Bodenhausen, G. V., Richeson, J. A., & Galinsky, A. D. (2011). Perspective taking combats automatic expressions of racial bias. *Journal of personality and social psychology*, *100*(6), 1027.

Tung, F.W. (2016). Child perception of humanoid robot appearance and behavior. *International Journal of Human-Computer Interaction*, *32*(6), 493-502,

Vinanzi, S., Patacchiola, M., Chella, A., & Cangelosi, A. (2019). Would a robot trust you? Developmental robotics model of trust and Theory of Mind. *Philosophical Transaction of the Royal Society B*, *374*(1771), 20180032.

Wainer, J., Robins, B., Amirabdollahian, F., & Dautenhahn, K. (2014). Using the humanoid robot KASPAR to autonomously play triadic games and facilitate collaborative play among children with autism. *IEEE Transactions on Autonomous Mental Development*, *6*(3), 183-199.

Wang, Y., Park, Y.-H., Itakura, S., Henderson, A. M. E., Kanda, T., Furuhata, N., & Ishiguro, H. (2019). Infants' perceptions of cooperation between a human and robot. *Infant and Child Development*, e2161.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*, 655-684.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103-128.

Woods, S. (2006). Exploring the design space of robots: Children's perspectives. *Interacting with Computers*, *18*(6), 1390-1418.

Woods, S., Dautenhahn, K., & Schulz, J. (2004). The design space of robots: Investigating children's views. In *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759)* (pp. 47-52). IEEE (September).

Yogeeswaran, K., Złotowski, J., Livingstone, M., Bartneck, C., Sumioka, H., & Ishiguro, H. (2016). The interactive effects of robot anthropomorphism and robot ability on perceived threat and support for robotics research. *Journal of Human-Robot Interaction*, *5*(2), 29-47.

Zanatto, D., Patacchiola, M., Cangelosi, A., & Goslin, J. (2019a). Generalisation of anthropomorphic stereotype. *International Journal of Social Robotics*, 1-10.

Zanatto, D., Patacchiola, M., Goslin, J., & Cangelosi, A. (2019b). Investigating cooperation with robotic peers. *PloS One*, *14*(11).

Zhang, T., Zhu, B., Lee, L., & Kaber, D. (2008). Service robot anthropomorphism

and interface design for emotion in human-robot interaction. In *2008 IEEE International Conference on Automation Science and Engineering* (pp. 674-679). IEEE (August).

Zhong, J., Ogata, T., Cangelosi, A., & Yang, C. (2019). Disentanglement in conceptual space during sensorimotor interaction. *Cognitive Computation and Systems*, *1*(4), 103-112.

Złotowski, J., Proudfoot, D., Yogeeswaran, K., & Bartneck, C. (2015). Anthropomorphism: Opportunities and challenges in human-robot interaction. *International Journal of Social Robotics*, *7*(3), 347-360.

7. Do We Take a Robot's Needs into Account?

The Effect of Humanization on Prosocial Considerations Toward Other Human Beings and Robots

S.R.R. Nijssen, E. Heyselaar, B.C.N. Müller, T. Bosse

ABSTRACT

Robots are becoming an integral part of society, yet the extent to which we are prosocial toward these nonliving objects is unclear. While previous research shows that we tend to take care of robots in high-risk, high-consequence situations, this has not been investigated in more day-to-day, low-consequence situations. Thus, we utilized an experimental paradigm (the Social Mindfulness 'SoMi' paradigm) that involved a trade-off between participants' own interests and their willingness to take their task partner's needs into account. In two experiments, we investigated whether participants would take the needs of a robotic task partner into account to the same extent as when the task partner was a human (Study I), and whether this was modulated by participant's anthropomorphic attributions to said robot (Study II). In Study I, participants were presented with a social decision-making task, which they performed once by themselves (solo context) and once with a task partner (either a human or a robot). Subsequently, in Study II, participants performed the same task, but this time with both a human and a robotic task partner. The task partners were introduced via neutral or anthropomorphic priming stories. Results indicate that the effect of humanizing a task partner indeed increases our tendency to take someone else's needs into account in a social decision-making task. However, this effect was only found for a human task partner, not for a robot. Thus, while anthropomorphizing a robot may lead us to save it when it is about to perish, it does not make us more socially considerate of it in day-to-day situations.

This chapter was originally published as Nijssen, S.R., Heyselaar, E., Müller, B.C., & Bosse, T. (2021). Do we take a robot's needs into account? The effect of humanization on prosocial considerations toward other human beings and robots. *Cyberpsychology, Behavior, and Social Networking, 24*(5), 332-336. We thank Sanne Derks, Arno Kok, and Moritz Sahay for their assistance with data collection. Preregistrations can be found here (https://osf.io/x2mcr/?view_only=fb0caac8a0694a96822a7ff2a1931ee0). No competing financial interests exist. This research was funded by Jacobs Foundation Grant 2014-1155. Supplementary Material.

Introduction

A total of 41.8 million robots are projected to be part of households around the world by 2020 (International Federation of Robotics, 2016). In to conduct tasks that are menial to humans, such as domestic assistance (Gross et al., 2015). While these technological developments can enrich people's lives (Riva et al., 2012) they also pose challenges. For example, the appearance of robots is becoming increasingly human-like, causing us to view them as more than tools (Roese & Amir, 2009). In situations where the life of a human-like robot is threatened, people sacrifice a group of anonymous humans to save that robot's 'life' (Nijssen et al., 2019) or empathize with a robot when it is physically mistreated (Riek et al., 2009).

Yet, these kinds of scenarios do not reflect our day-to-day interactions with robotic devices. On a day-to-day basis, the majority of our prosocial behavior toward other humans consists of small acts of prosociality (e.g., giving someone directions or considering someone's perspective when taking a decision). Clearly, not wanting someone or something to perish involves different affective motivations than giving directions. Indeed, prosocial behavior increases as a function of urgency (Christensen et al., 1998) and potential harm (Shotland & Stebbins, 1983). Similarly, people empathize more with a robot that is being severely maltreated compared to a robot that is being treated kindly (Rosenthal-von der Pütten et al., 2013). Thus, while previous research shows that we can be moral or empathetic toward robots in urgent and high-consequence situations, we do not know whether day-to-day, low-consequence acts of prosocial behavior occur as well. Yet, since robotic devices are becoming increasingly common in the household, it is pertinent to understand the mechanisms that ground our common, day-to-day interactions with them. Are we prosocial toward robotic devices, and if so, which factors influence this?

Prosocial behavior consists of actions that benefit people other than oneself, such as helping, sharing, or comforting (Batson, 1998). In human-human interaction, prosocial behavior allows us to form and maintain relationships with people. For us to behave in a way that benefits another individual, we need to be able to consider that individual's needs and desires. Thus, prosocial behavior involves the ability to take another person's perspective (Batson et al., 1997), understand that their needs and desires may be different from our own (Baron-Cohen et al., 1994), and possibly experience empathic concern toward them (Bateson, 1987; Zahn-Waxler & Radke-Yarrow, 1990).

Prosocial behavior presupposes that the target of our prosocial action *has* mental states and *has* affective experiences. Arguably, currently

available robots do not have emotional experiences or needs similar to humans (Davies, 2007). Yet, humans have a strong tendency to perceive *nonhuman agents* in human terms by attributing mental states, emotions, and intentions to them: a phenomenon called anthropomorphism (Epley et al., 2007; Heider & Simmel, 1944). Specifically relating to the human-machine interaction, the Media Equation theory (Reeves & Nass, 1996) and Computers As Social Actors (CASA) paradigm (Nass & Moon, 200) claim that humans will respond to any type of object (such as a computer) as if they are human, provided enough social cues are present.

The link between prosociality and the attribution of mental states to nonhuman agents opens up the intriguing possibility that we act prosocially toward robots when we anthropomorphize them. Indeed, anthropomorphizing a nonhuman agent has been shown to lead to more interpersonal connectedness (Müller et al., 2014) and higher trust (Waytz et al., 2014). Specifically, anthropomorphizing a robot leads to processing its movement as human (Stenzel et al., 2012), joint attention (Wykowska et al., 2014), increased trust (Eyssel et al., 2012), and moral care (Nijssen et al., 2019; Riek et al., 2009; Rosenthal-von der Pütten et al., 2013). However, previous research on human-robot interaction was done in urgent and high-consequence situations. Therefore, results cannot be generalized to low-consequence situations.

An established paradigm for measuring everyday, low-consequence acts of prosociality is the Social Mindfulness (SoMi) paradigm (Van Doesum et al., 2013). The SoMi paradigm measures our willingness to consider another individual's needs before our own. In the SoMi task, participants have to repeatedly choose among three items of the same category (e.g., pens). Crucially, two items are identical, while one item differs in a certain aspect (e.g., two blue and one black pen). Participants are told that they have to choose an item, but that someone else will pick something from the remaining items after them. It is counted how often the participant picks the socially mindful item (of which there were two so the task partner still has a choice between two unique items). The overall proportion of socially mindful versus nonsocially mindful choices thus gives an indication of a participant's overall willingness to consider the task partner's needs. In human-human interaction, participants have shown to be more socially mindful in their decision-making when another individual has to pick after them (Van Doesum, 2018). Furthermore, SoMi scores have been correlated with measures such as general empathy (Van Doesum et al., 2013).

The current research investigated social mindfulness toward robots. Two separate studies were conducted utilizing the SoMi paradigm. In Study I, participants were presented with two experimental blocks: in

one, they performed the SoMi task alone, that is, they had to make choices between items without a partner. In the other block, they performed the classic SoMi task with a partner (a human or a robot). We hypothesized that (1a) more socially mindful choices would be made in the social condition than in the solo condition; and we further expected that (1b) participants would make more socially mindful decisions when the partner was another human compared with a robot.

In Study II, we investigated the effect of anthropomorphic attributions on the level of social mindfulness. Participants performed the SoMi task with another human and a robot, both of which were described in either human-like and anthropomorphic terms or in neutral and mechanical terms. We hypothesized (2a) participants in the anthropomorphic condition to be more socially mindful than participants in the neutral condition, (2b) regardless of whether their partner was a human or a robot.

Study I

Methods

Participants. The minimum required sample was determined to be $n=134$ (67 participants per condition), using G*Power (with $\alpha=0.05$, $\beta=0.80$, $dz=0.35$). The effect size used was based on a previous effect size comparing responses with humans and robots (Nijssen et al., 2019) since this study used a similar experimental design. This effect size was divided by two to obtain a conservative estimate.

A total of $n=136$ undergraduate students were recruited to participate in this study in exchange for course credit. All provided informed consent before participation. Based on preregistered exclusion criteria, 12 participants were dropped from the analysis for completing the study in less than 3 minutes, as well as 2 participants who took more than 90 minutes. This resulted in a final sample of 122 students ($M_{age}=22.11\pm4.98$, 83 females).

Materials and procedure. Participants completed the experiment on their own computer in a quiet environment using Qualtrics. After signing up, participants received a link to the online experiment. Before the experiment started, participants were instructed to ensure they would not be interrupted for the duration of the experiment (±20 minutes).

The solo condition was completed by all participants. Participants were randomly assigned to the social-human or social-robot condition. The presentation order of the solo and social block was counter-

balanced. Each block consisted of 12 trials (6 test- and 6 distractor trials). During test trials, participants were presented with three items from which they had to choose one. In all test trials, two objects were identical and one object was different. The distractor trials contained four similar items: two of each type. The presentation order of the test and distractor trials and the order of items were randomized.

In the solo condition, participants were instructed to imagine that they can take the object they chose home with them. In the social condition, participants were informed that someone else would choose between the remaining items. In both conditions, a picture of the task partner was displayed (Nijssen et al., 2019; Langner et al., 2010). Subsequently, participants provided their age and gender, were thanked, debriefed, and awarded course credit. The experimental procedure was approved by the institutional review board of the affiliated university. Specific examples of our materials as well as details of the data analysis procedure can be found in the Supplementary Material[1].

Results

Participants' choices in the social mindfulness trials were analyzed using a binomial mixed-effects logit model (Bates et al., 2014) in R. No significant effect of task partner was found, $p=0.697$. The proportion of socially mindful choices did not differ between the solo ($M=0.53\pm0.20$) and the social-human ($M=0.55\pm0.23$) or social-robot ($M=0.56\pm0.25$) conditions.

Discussion

Contrary to hypotheses 1a and 1b, we could not detect any differences between the solo and social conditions, nor between the human and robot task partners in the social condition.

Looking at the proportions of socially mindful choices, our results align with other 'neutral' conditions in the SoMi paradigm (Van Doesum et al., 2018; Van Doesum et al., 2013). Similar to our findings, Van Doesum et al. (2013) report social mindfulness proportions at chance level when participants are not given any explicit instructions; only after explicit instruction to take their partner's perspective do they find a significant increase in social mindfulness. In addition, in a real-life ver-

[1] https://www.liebertpub.com/doi/suppl/10.1089/cyber.2020.0035/suppl_file/Supp_Data.zip.

sion of the task, Van Doesum et al. (2018) report a chance level of social mindfulness when no partner is present. However, when a partner is present, the rate of social mindfulness increases significantly.

Comparing this with our findings, we can draw two preliminary conclusions. First, the level of social mindfulness in our solo condition is on par with similar human-human baseline conditions in the literature. Second, our social condition did not sufficiently trigger participants to take the perspective of their partner as we only presented a picture and provided no other information. As previous research used explicit perspective-taking instructions, we thus designed a follow-up study in which participants were induced to take their partner's perspective using vignettes. Participants' social mindfulness toward a robotic and human task partner was assessed, while vignettes of the task partners were presented to induce participants' attributions of mental states.

Study II

Methods

Participants. The minimum required sample size was $n=128$ (64 per group), using G*Power ($\alpha=0.05$, $\beta=0.80$, $\eta_p^2=0.059$). A total of $n=128$ ($M_{age}=26.54\pm11.10$, 78 females) participants participated in the study in exchange for course credit. Participants were recruited both in the participant pool of the researchers' local university and through social media. All participants provided informed consent before their participation.

Materials and procedure. This experiment had a 2 (Partner: human vs. robot; within-subjects) by 2 (Condition: anthropomorphic vs. neutral; between-subjects) mixed design. Participants were randomly assigned to conditions. The materials and procedure were similar to Study I except that no solo condition was present. The presentation order of the human and robot block was counterbalanced.

Each block consisted of an introduction to the task partner followed by 12 SoMi trials. Participants were introduced to their task partner with a vignette. In the anthropomorphic condition, these vignettes described the task partner in a humanized manner, that is, by emphasizing mental states such as their emotions and intentions. In the neutral condition, vignettes described the task partner in a neutral manner, without referring to their mental states (for more details about the vignettes, see Majdandžićet al., 2016; Nijssen et al., 2019). A picture of the task partner was presented next to the vignettes. The combination of the vignettes

and pictures was counterbalanced. The task partner was consistently referred to with a letter (e.g., 'H'). Importantly, participants in the anthropomorphic condition read a humanized vignette for both the robotic partner and the human partner; idem for participants in the neutral condition. The humanizing versus neutral manipulation effect of the vignettes was validated in previous work (Majdandžić et al., 2016; Nijssen et al., 2019). Subsequently, participants provided their age and gender, were thanked, debriefed, and awarded course credit. The experimental procedure was approved by the institutional review board of the affiliated university. Specific examples of our materials as well as details of the data analysis procedure can be found in the Supplementary Material[2].

Results

Participants' choices in the social mindfulness trials were analyzed using a binomial mixed-effects logit model (Bates et al., 2014) in R. A significant main effect of condition was found ($\beta = -0.29$, Wald $Z = -2.95$, $p < 0.001$), with a larger proportion of socially mindful choices made in the anthropomorphic ($M = 0.64 \pm 0.19$) versus the neutral ($M = 0.52 \pm 0.20$) condition. No significant main effect was found for task partner ($\beta = 0.10$, Wald $Z = 1.58$, $p = 0.087$). However, a significant interaction effect of condition and task partner ($\beta = -0.14$, Wald $Z = -2.19$, $p = 0.017$) was found.

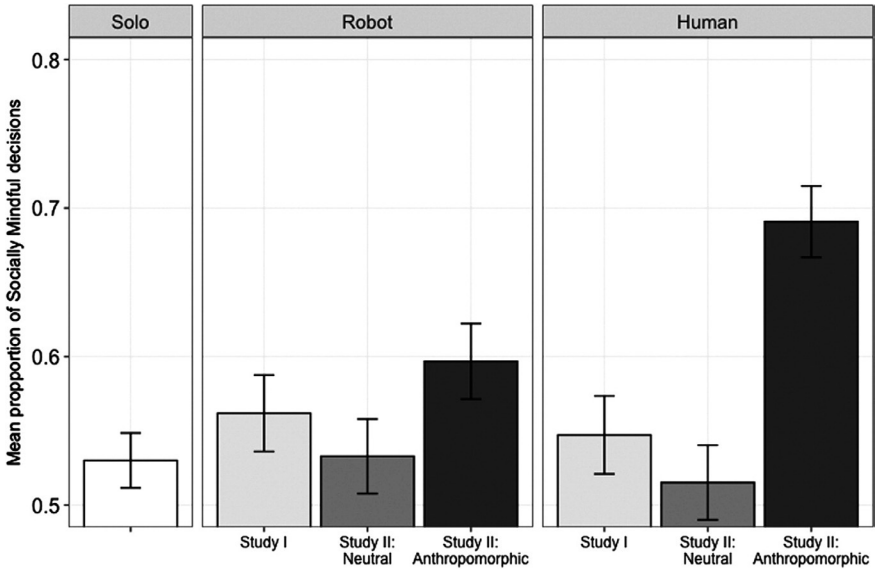
Participants made significantly more socially mindful decisions toward the human partner in the anthropomorphic ($M = 0.69 \pm 0.23$) versus the neutral ($M = 0.52 \pm 0.24$) condition ($\chi^2 = 24.68$, $p < 0.001$). For the robot partner, this difference did not reach statistical significance ($M = 0.60 \pm 0.24$ in the anthropomorphic condition vs. $M = 0.53 \pm 0.27$ in the neutral condition; $p = 0.074$). The significant interaction thus indicates that the difference in social mindfulness between the neutral and anthropomorphic conditions was significantly higher for the human partner than the robot partner.

Two additional binomial mixed-effects logit models with the social condition (picture-only) from Study I as the reference group (Figure 1) showed no significant difference between conditions for the robot partner (all ps > 0.396). For the human partner, the anthropomorphic condition yielded significantly more socially mindful choices than the social (picture-only) condition from Study I ($\beta = 0.72$, Wald $Z = 3.53$, $p < 0.001$). The neutral condition from Study II and social (picture-only) condi-

[2] https://www.liebertpub.com/doi/suppl/10.1089/cyber.2020.0035/suppl_file/Supp_Data.zip.

tion from Study I did not statistically differ (β=–0.14, Wald Z=–0.74, p=0.338).

Figure 1 - *Overview of statistical results*



Note: In the *left-hand panel*, the proportion of socially mindful choices in the solo condition of Study I is displayed. In the *middle panel*, the proportion of socially mindful choices with a robot task partner is illustrated, in the picture-only condition of Study I as well as in the neutral and anthropomorphic conditions of Study II. The same conditions are displayed for the human task partner in the *right-hand panel*.

General Discussion

The current research investigated whether anthropomorphizing robots would affect prosocial behavior in a social decision-making task. Results of Study I show that the same level of socially mindful decisions was made in a solo context as with a human or robot partner. Results of Study II show that participants became significantly more socially mindful in the anthropomorphic condition, thus confirming hypothesis 2a. However, this effect was only found for the human partner: participants took the needs of their human partner into account more often when their mental states and emotions were emphasized than when they were not. For the robot partner, the difference in level of social mindfulness

between the neutral and anthropomorphic condition was not significant – thus rejecting hypothesis 2b.

This research utilized an experimental paradigm that allowed us to measure day-to-day, low-consequence prosocial behavior. This stands in sharp contrast to previous research on prosocial behavior toward robots, which relied on urgent and high-consequence scenarios. While previous research showed that people in such scenarios are indeed more likely to, for example, protect an anthropomorphized robot from harm (Nijssen et al., 2019), trust an anthropomorphized car more despite it causing an accident (Waytz, Heafner & Epley, 2014), or empathize more with a robot that is being physically mistreated (Riek et al., 2009), our results indicate that this effect of anthropomorphism does not necessarily extend to more common, low-consequence prosocial considerations. This is in line with previous research showing that participants empathize more with a suffering than a nonsuffering robot (Rosenthal-von der Pütten et al., 2013). Moreover, our results match the dynamics of prosocial behavior in human-human interaction: people tend to be more prosocial toward someone when the urgency (Christensen et al., 1998) or potential harm (Shotland & Stebbins, 1983) of their situation increases. Linking this to our findings, the social decision-making trials in our two experiments were not constructed as highly urgent or highly harmful situations. The task partner simply lost out on an opportunity to choose between everyday items. In sum, our findings point to a relevant distinction based on urgency in the effects of anthropomorphism for human-robot interaction.

It should be noted, however, that the neutral vignettes used in this study also included some anthropomorphic cues regarding a robot's autonomy. Moreover, it could be argued that the content of the vignettes confounded participants' perception of the agent. However, the distinct effect of the neutral versus humanizing vignettes on mentalizing, liking, and empathy was confirmed in previous research (Majdandžić et al., 2016; Nijssen et al., 2019). In addition, previous studies used several different vignettes for each category (humanizing vs. neutral) and found no differences between the individual vignettes in each category; furthermore, the vignettes that were quantitatively confirmed in previous studies as most distinctly emphasizing humanness versus neutrality were selected for the current study. Even though we thus confirmed the humanizing versus neutral effects of the manipulation in the current study, it would be interesting to include measures of empathy and liking as potential mediators in follow-up research.

This research is among the few investigations of human-robot interaction that directly compared participants' behavior toward robots with behavior toward humans. Many experiments on human-robot interaction focus solely on how anthropomorphizing versus nonanthropomor-

phizing affects certain behavioral parameters (Eyssel et al., 2012). However, our results show a clear distinction in the effects of our manipulation on human versus robot partners. Thus, if studies on human-robot interaction want to investigate certain behavioral parameters and draw conclusions about how those behaviors compare with human-human interaction, including an experimental condition in which those parameters are measured vis-à-vis another human seems pertinent.

While our findings are relevant and have clear theoretical and practical implications, they are not in line with the long-standing Media Equation Theory and the CASA paradigm (Nass & Moon, 2000; Reeves & Nass, 1996). While these models of human-machine interaction entail that we treat any machine as a human as long as it displays sufficient social cues, in our study, anthropomorphic vignettes did not significantly increase participants' social mindfulness toward robots. This may be because our study did not involve any real-life interactions, in contrast to the empirical studies by Nass and Moon (2000). Our results should thus be corroborated in future research, especially in real-life human-robot interaction settings.

Given the increasing role of robotic devices in our daily lives, a better understanding of the mechanisms that support our interactions with them is pertinent. Our results show that effects of anthropomorphism cannot be generalized across different types of social interactions: while we may be inclined to care for an anthropomorphic robot when it is about to be demolished does not mean we take its needs and desires into account in an ordinary situation.

References

Baron-Cohen, S., Tager-Flusberg H., & Cohen, D. J. (Eds.). (1994). *Understanding Other Minds: Perspectives from Autism. Understanding Other Minds.* Oxford University Press.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1-48.

Batson C. (1998) Altruism and prosocial behavior. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *The Handbook of Social Psychology* (pp. 282-316). McGraw-Hill.

Batson, C. D. (1987). Prosocial motivation: Is it ever truly altruistic?. *Advances in experimental social psychology, 20*, 65-122.

Batson, C. D., Early, S., & Salvarani, G. (1997). Perspective taking: Imagining how another feels versus imaging how you would feel. *Personality and social psychology bulletin, 23*(7), 751-758.

Christensen, C., Fierst, D., Jodocy, A., & Lorenz, D. N. (1998). Answering the call for prosocial behavior. *The Journal of social psychology, 138*(5), 564-571.

Davies, M. Y. A. (2007). Taking robots personally: A personalist critique. In T. Metzler (Ed.), *Human Implications of Human-Robot Interaction: Technical Report WS-07-07* (pp. 5-8). AAAI.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological review, 114*(4), 864.

Eyssel, F., De Ruiter, L., Kuchenbrandt, D., Bobinger, S., & Hegel, F. (2012, March). 'If you sound like me, you must be more human': On the interplay of robot and user features on human-robot acceptance and anthropomorphism. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 125-126). IEEE.

Gross, H. M., Mueller, S., Schroeter, C., Volkhardt, M., Scheidig, A., Debes, K., Richter, K., & Doering, N. (2015). Robot companion for domestic health assistance: Implementation, test and case study under everyday conditions in private apartments. In *Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 5992-5999).

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American journal of psychology, 57*(2), 243-259.

International Federation of Robotics. (2016). Distribution of service robots. In T. Visti (Ed.), *2016 World Robotics Report* (pp. 18-29). International Federation of Robotics.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & Van Knippenberg, A. D. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and emotion, 24*(8), 1377-1388.

Majdandžić, J., Amashaufer, S., Hummer, A., Windischberger, C., & Lamm, C. (2016). The selfless mind: How prefrontal involvement in mentalizing with similar and dissimilar others shapes empathy and prosocial behavior. *Cognition, 157*, 24-38.

Müller, B. C., van Baaren, R. B., van Someren, D. H., & Dijksterhuis, A. (2014). A present for Pinocchio: On when non-biological agents become real. *Social Cognition, 32*(4), 381-396.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of social issues, 56*(1), 81-103.

Nijssen, S. R., Müller, B. C., Baaren, R. B. V., & Paulus, M. (2019). Saving the robot or the human? Robots who feel deserve moral care. *Social Cognition, 37*(1), 41-S2.

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

Riek, L. D., Rabinowitch, T. C., Chakrabarti, B., & Robinson, P. (2009). How anthropomorphism affects empathy toward robots. In *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction* (pp. 245-246). IEEE.

Riva, G., Baños, R. M., Botella, C., Wiederhold, B. K., & Gaggioli, A. (2012). Positive technology: Using interactive technologies to promote positive functioning. *Cyberpsychology, Behavior, and Social Networking, 15*(2), 69-77.

Roese, N. J., & Amir, E. (2009). Human-android interaction in the near and distant future. *Perspectives on Psychological Science, 4*(4), 429-434.

Rosenthal-von der Pütten, A. M., Krämer, N. C., Hoffmann, L., Sobieraj, S., & Eimler, S. C. (2013). An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics, 5*(1), 17-34.

Shotland, R. L., & Stebbins, C. A. (1983). Emergency and cost as determinants of helping behavior and the slow accumulation of social psychological knowledge. *Social psychology quarterly, 46*(1), 36-46.

Stenzel, A., Chinellato, E., Bou, M. A. T., Del Pobil, Á. P., Lappe, M., & Liepelt, R. (2012). When humanoid robots become human-like interaction partners: Corepresentation of robotic actions. *Journal of Experimental Psychology: Human Perception and Performance, 38*(5), 1073.

Van Doesum, N. J., Karremans, J. C., Fikke, R. C., de Lange, M. A., & Van Lange, P. A. (2018). Social mindfulness in the real world: the physical presence of others induces other-regarding motivation. *Social Influence, 13*(4), 209-222.

Van Doesum, N. J., Van Lange, D. A., & Van Lange, P. A. (2013). Social mindfulness: Skill and will to navigate the social world. *Journal of personality and social psychology, 105*(1), 86.

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology, 52*, 113-117.

Wykowska, A., Wiese, E., Prosser, A., & Müller, H. J. (2014). Beliefs about the minds of others influence how we process sensory information. *PloS One, 9*(4), e94339.

Zahn-Waxler, C., & Radke-Yarrow, M. (1990). The origins of empathic concern. *Motivation and emotion, 14*(2), 107-130.

8. Robots Are Not All the Same

Young Adults' Expectations, Attitudes, and Mental Attribution to Two Humanoid Social Robots

F. Manzi, D. Massaro, D. Di Lernia, M.A. Maggioni, G. Riva, A. Marchetti

ABSTRACT

The human physical resemblance of humanoid social robots (HSRs) has proven to be particularly effective in interactions with humans in different contexts. In particular, two main factors affect the quality of human-robot interaction, the physical appearance and the behaviors performed by the robot. In this study, we examined the psychological effect of two HSRs, NAO and Pepper. Although some studies have shown that these two robots are very similar in terms of the human likeness, other evidence has shown some differences in their design affecting different psychological elements of the human partner. The present study aims to analyze the variability of the attributions of mental states (AMS), expectations of robotic development and negative attitudes as a function of the physical appearance of two HSRs after observing a real interaction with a human (an experimenter). For this purpose, two groups of young adults were recruited, one for the NAO (N=100, M=20.22) and the other for the Pepper (N=74, M=21.76). The results showed that both the observation of interaction and the type of robot affect the AMS, with a greater AMS to Pepper robot compared to NAO. People's expectations, instead, are influenced by the interaction and are independent of the type of robot. Finally, negative attitudes are independent of both the interaction and the type of robot. The study showed that also subtle differences in the physical appearance of HSRs have significant effects on how humans perceived robots.

Introduction

In the last 20 years, different types of humanoid social robots (HSRs) were built to help and support people in different contexts from domes-

---

This chapter was originally published as Manzi, F., Massaro, D., Di Lernia, D., Maggioni, M.A., Riva, G., & Marchetti, A. (2021). Robots are not all the same: Young adults' expectations, attitudes, and mental attribution to two humanoid social robots. *Cyberpsychology, Behavior, and Social Networking, 24*(5), 307-314. No competing financial interests exist. This research was funded by Università Cattolica del Sacro Cuore (Human-Robot Confluence project).

tic to health care to educational (Belpaeme et al., 2018; Marchetti et al., 2018; Marchetti et al., 2020a,b). In these contexts, the physical human likeness of the HSRs has proven to be particularly effective in the interactions with humans (Belpaeme et al., 2018; Dario et al., 2001; Robins et al., 2006; Wu et al., 2012). A recent review of the literature by Marchetti et al. (2018) on humanoid robots showed that the quality of human-robot interactions (HRIs) improves as a function of at least two factors: the physical appearance of the robot (i.e., the design) and the behaviors performed by the robot (i.e., the interactional routines).

In the present study, the two HSRs that are used, NAO and Pepper, differ in terms of their physical appearance – although as shown below this difference is subtle –, while the behaviors performed by the HSRs are the same. Therefore, we intend here to analyze the variability of attributions of mental states (AMS), robotic development expectations, and negative attitudes as a function of the physical appearance of NAO and Pepper after observing a real interaction with a human (an experimenter).

The design of the HSRs: NAO and Pepper

The study of the physical design of HSRs is an important research topic in the field of HRI studies, although it has not yet been systematically analyzed from a psychological point of view. Several studies have shown that specific facial features (i.e., nose, eyes, and mouth) resembling human ones are particularly important for the perception of the robots' human-likeness (DiSalvo et al., 2002; Duffy, 2003; Manzi et al., 2020b; Prakash & Rogers, 2015). However, excessive resemblance to the human triggers the Uncanny Valley effect: the more the appearance of robots is similar to humans, the more is the sense of eeriness (MacDorman & Ishiguro, 2006; Mori, 1970; Mori et al., 2012). The human sensitivity to the physical appearance of robots was also found in childhood (Di Dio et al., 2020a,b,c; Manzi et al. 2020a), in particular, children already at 5 years of age attribute greater mental qualities to a more anthropomorphic robot, that is, the NAO, compared to a more mechanical robot, that is, the Robovie (Manzi et al. 2020b).

A recent study classified around 200 HSRs based on their level of human likeness (Phillips et al., 2018). In this collection, the NAO robot was rated on the human-like factor with a score of 46/100, while the Pepper robot with a score of 42/100. Although these two HSRs are very similar with respect to their level of human likeness, they nevertheless display physical differences both substantial (e.g., NAO is shorter than the Pepper and NAO has legs, while the Pepper has a platform with wheels) and

more nuanced (e.g., Pepper has bigger eyes and has five fingers, while NAO has less defined eyes and only three fingers).

Furthermore, Pandey & Gelin (2018) stressed that the voice implemented in the two robots is different to overcome some stereotypes and unrealistic expectations toward them. The authors reported that, for the NAO, people generally use the pronoun *him*, while for the Pepper they use *him*, *her*, or *it* almost equally. This denotes greater gender neutrality for the robot Pepper.

In support of the hypothesis of the presence of differences between these two robots, despite the fact that they are very similar HSRs, a recent study by Thunberg et al. (2017) showed that people after experiencing a situation where the NAO robot and Pepper robot opposing to an experimenter's request, participants judged the NAO as more likable, intelligent, safe, and lifelike than Pepper. The authors suggested that these differences could result from a more pronounced social presence of Pepper – due to its physical appearance – compared to the NAO and, consequently, participants attributed lower positive social qualities to the Pepper robot.

In this sense, although some studies show that these two robots are very similar HSRs, other evidence indicates that their physical appearance may exert different effects on human's perception of their psychological attributes.

The effect of interactions

In psychology, it is widely recognized that from early childhood, representations of relationships change over time because they are conditioned by experiences with others (Ainsworth, 1969; Ainsworth et al., 1971; Bowlby, 1969, 1973, 1980). As a matter of fact, people tend to change their opinions and attitudes about others as a result of their reciprocal experiences.

As for human interactions, although with a different relational meaningfulness, experiencing an interaction with a robot could change the representation that a person has of it. Several studies analyzed this variation in terms of social presence, pleasantness, and, more generally, attribution of psychological qualities, showing that even the mere observation of the interaction between a human and a robot increases the perception of human characteristics of the robot and vice-versa decreases the negative psychological contents in humans (Beuscher et al., 2017). In particular, the higher level of human likeness of the robot positively influences the attribution of human-like qualities to the robot (Manzi et al., 2020b; Zanatto et al., 2020). Therefore, the positive effect of interaction also depends on the type of robot used.

In a recent study with the NAO robot, it was found that negative attitudes toward the robot, specifically anxiety, decreased significantly after experiencing a single interaction, particularly in elderly compared to a group of high-school students (Sinnema & Alimardani, 2019). In addition, another study with the Pepper robot found that adults increase their perception of the robot's social presence after one interaction (Edwards et al., 2019). Also, a study by Thunberg et al. (2017) comparing NAO and Pepper showed that the negative attitudes toward these two robots do not differ after an interaction. However, the findings revealed that participants reported greater positive feelings toward NAO than Pepper. In general, we can state that interacting or observing an interaction improves the perception of human likeness of the robot and that depending on the type of robot the psychological effects are different.

Although several studies have shown that experiencing interactions with robots affects how humans perceive them in terms of psychological attributes (Beuscher et al., 2017; Edwards et al., 2019; Sinnema & Alimardani, 2019), no study has yet examined how people's expectations of robots in terms of technical and human properties change as a function of the interaction. In this sense, a recent study showed that people's expectations for the future development of robots can be classified according to five specific profiles that are placed on different levels of a robot humanization continuum. However, this study did not analyze the expectations after observing an interaction.

Aims of the study

The present study aimed to investigate the AMS, people's expectations, and negative attitudes in young adults toward two HSRs, NAO and Pepper, varying in their physical appearance (DiSalvo et al., 2002; Duffy, 2003; Phillips et al., 2018). Differences in the attribution of mental qualities, people's expectations and negative attitudes toward the two robots were then explored as function of the type of robot and the observation of an interaction between a human and the robots.

Methods

Participants

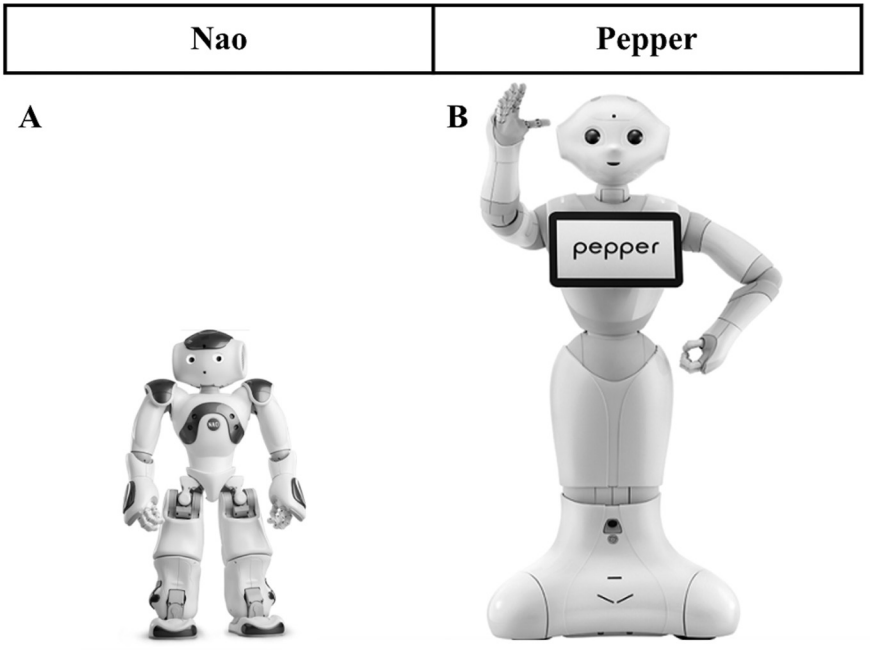
Data were acquired on 174 Italian university students from humanistic degree courses. The participants were divided by the two interaction conditions with the Pepper robot (N=100, M=20.22, SD=1.80, F=90)

and the NAO robot (N=74, M=21.76, SD=4.42, F=41). Each subject gave their written consent to participate in the study.

Description of robots

NAO robot. The V6 edition of the NAO (SoftBank Robotics®) robot (Figure 1A) was used. NAO is a bipedal robot with pleasantly rounded features, about 58 cm high. It has 25 degrees of freedom to move and adapt to the environment. NAO is a humanoid robot with pronounced anthropomorphic characteristics: two legs, two arms, and two hands with three moving fingers. Moreover, the face is composed of two eyes and a camera placed in the lower part of the face that resembles a mouth. Overall, it looks like a child. In addition, it performs quite fluid actions, similar to the movements of a child.

Figure 1 (A, B) - *Images representing the robot NAO and the robot Pepper*



Pepper robot. The Pepper robot (SoftBank Robotics) was used (Figure 1B). Pepper is 120 cm high and its design guarantees a high level of user acceptance. It has 20 degrees of freedom for natural and expressive movements. The Pepper compared to the NAO does not have legs,

but a platform with wheels that allow it to move in the environment. Despite the lack of legs, Pepper seems highly anthropomorphic, especially due to the two arms and hands with five moving fingers. Moreover, the face has big eyes (the video cameras) that make it look very similar to a person; furthermore, the camera in the lower part of the face seems a mouth. In general, Pepper looks like an adult robot.

Measures

Attribution of mental states. The AMS (Manzi et al., 2020b) is a measure of the attribution of mental qualities that participants ascribe when they look at images depicting specific characters (humans or robots) (Di Dio et al., 2020a,c). In the present study, according to the group condition, participants AMS to the NAO or the Pepper robot. This scale is a self-administer instrument composed by 25 items evaluated on a 10-point Likert scale (1 = not at all; 10 = absolutely yes).

The scale consists of five mental states dimensions: the first dimension measures 'Perceptive' states (five items; e.g., 'Do you think it can smell?'); the second dimension concerns the 'Emotive' states (five items; e.g., 'Do you think it can be happy?'); the third dimension measures 'Intentions and Desires' states (five items; e.g., 'Do you think it can make a wish?'); the fourth dimension concerns 'Imaginative' states (five items; e.g., 'Do you think it can make a joke?'); and the fifth dimension measures 'Epistemic' states (five items; e.g., 'Do you think it can teach?'). All the dimensions resulted to be reliable (respectively: preinteraction $\alpha = 0.78$, 0.95, 0.85, 0.82, 0.83; postinteraction $\alpha = 0.83$, 0.97, 0.91, 0.87, 0.74).

Scale for robotic needs. The Scale for Robotic Needs (SRN) (Manzi et al., 2021) evaluates participants' expectations about the development of the robots, in particular, those characteristics/skills that the robot should acquire to be more both technical and relational efficient. This scale is a self-administer instrument composed by 17 items evaluated on a 5-point Likert scale (1 = not at all; 5 = very much).

The SRN's factorial structure consists of four factors: three items measure expectation about robots' 'Technical Features' [e.g., 'Have more power (e.g., battery life)'; pre-interaction $\alpha = 0.64$, post-interaction $\alpha = 0.78$], three items measure expectation about their 'Social-Emotional Resonance' (e.g., 'Increase its ability to transmit and communicate emotions'; pre-interaction $\alpha = 0.85$, post-interaction $\alpha = 0.76$), three items measure expectation about their 'Agency' (e.g., 'Have the ability to perform tasks with other people or robots'; pre-interaction $\alpha = 0.71$, post-interaction $\alpha = 0.75$), and the remaining eight items measure ex-

pectation about robots' 'Human Life' (e.g., 'Have a pet'; pre-interaction $\alpha = 0.89$, post-interaction $\alpha = 0.87$).

Negative attitudes toward robots scale. The NARS (Nomura et al., 2006) measures humans' attitudes toward robots in daily life. The scale consists of 14 items divided into three subscales: the first subscale measures 'Negative attitudes toward situations and interactions with robots' (six items; e.g., 'I would feel uneasy if I was given a job where I had to use robots'); the second subscale concerns the 'Negative attitudes toward social influence of robots' (five items; e.g., 'I would feel uneasy if robots really had emotions'); and the last subscale measures 'Negative attitudes toward emotion in interaction with robots' (three items to reverse; e.g., 'I feel comforted being with robots that have emotions'). Each item is evaluated on a 5-point Likert scale (1 = strongly disagree; 7 = strongly agree). All the subscales resulted to be reliable (respectively: pre-interaction $\alpha = 0.73$, 0.65, 0.68; post-interaction $\alpha = 0.76$, 0.77, 0.64).

Procedure

The experiment consisted of three phases. In the first phase, the 'Scale for Robotic Needs,' the 'Attribution of Mental States,' and 'Negative Attitudes towards Robots' were administered randomly. During the interaction phase, the experimenter interacted with NAO or Pepper to show the robot's speech recognition, response, and movement capabilities. Also, the human and the robot played a game named 'guess what it is': the robot explained the rules of the game, showing the experimenter a series of cards representing animals or fruits. The experimenter chose one of these cards and the robot, asking a series of questions, guessed which animal or fruits are represented on the card. The short game was designed to introduce the subjects also some 'autonomous' mental abilities of the robot simulating a cognitive and relational functioning. During the last phase of the experiment, the same tests of the pretest phase were administered randomly.

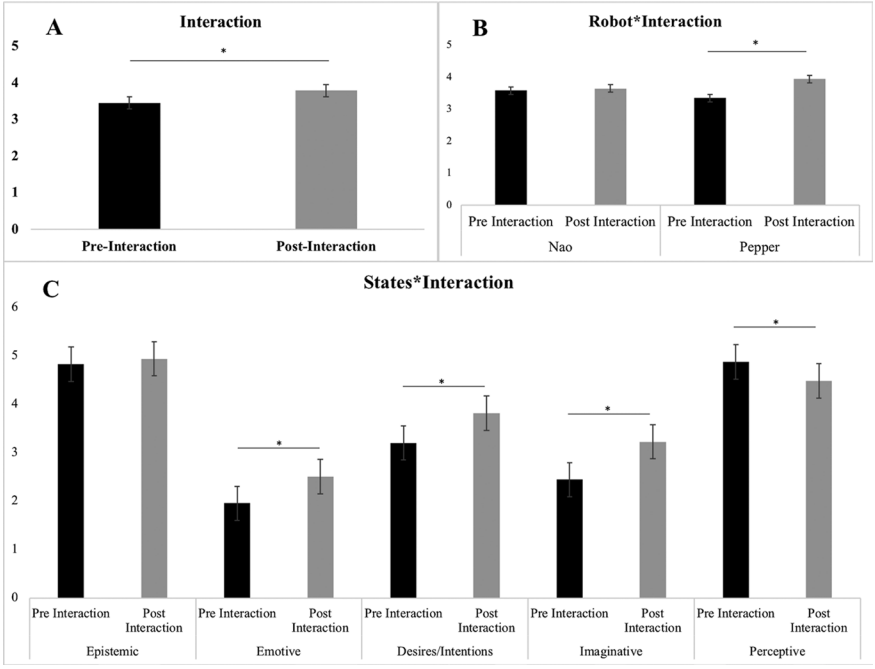
Results

To estimate the effect of observing an interaction between humans and robots on the AMS to the robots, on expectations of development of the future robots and on negative attitudes toward the robots, three generalized linear model (GLM) analyses were carried out. The Greenhouse-Geisser correction was used for violations of Mauchly's Test of Sphericity ($p < 0.05$). *Post hoc* comparisons were Bonferroni corrected.

The effect of the interaction on the AMS

To evaluate the effect of the type of robot on the AMS to the robots a GLM analyses was carried out with two levels of *Interaction* (pre-interaction, post-interaction) and five levels of *States* (epistemic, emotive, intention/desires, imaginative, perceptive) as within-subject factors and *Robot* (NAO, Pepper) as between-subject factor. The GLM revealed (1) a significant main effect of *Interaction* ($F=12.56$, $p=0.001$), suggesting that the interaction affect positively the AMS (post-interaction > pre-interaction, $M\text{Diff}=0.332$, $SE=0.094$, $p<0.001$; Figure 2A); (2) a significant main effect of *States* ($F=164.98$, $p<0.001$), suggesting that participants ascribed less emotional states compared to the other dimensions (Table 1).

Figure 2 (A-C) - *Bar chart representing: (A) the AMS mean scores as a function of Interaction; (B) the AMS mean scores as a function of Robot and Interaction; (C) the AMS mean scores as a function of States and Interaction (pre-interaction, post-interaction)*



Note: The bars represent the standard errors of the mean; *Indicates significant differences. AMS, attribution of mental states.

Table 1 - *Statistics comparing the attribution of epistemic and emotive states of the AMS and the other dimensions of the AMS*

AMS dimensions	*Mean Diff.*	*SE*	*Sign.b*
Epistemic			
Emotive	2.653	0.105	0.001
Desires/intentions	1.377	0.099	0.001
Imaginative	2.050	0.114	0.001
Perceptive	0.209	0.149	1
Emotive			
Epistemic	–2.653	0.105	0.001
Desires/intentions	–1.275	0.104	0.001
Imaginative	–0.603	0.093	0.001
Perceptive	–2.444	0.157	0.001

Note: AMS, Attribution of Mental States.

Furthermore, the analysis showed a two-way interaction between (1) *Robot* and *Interaction* (F=26.28, p=0.006; Figure 2B), suggesting that Pepper has a greater effect compared to NAO, and (2) *Condition* and *States* (F=17.21, p<0.001; Figure 2C), showing that all dimensions, except epistemic states, change after interaction. In addition, the results indicated a significant three-way interaction between *Robot* × *Interaction* × *States*, $F(1, 154) = 2.56$, p=0.037, *partial*-η^2=0.016, δ=0.72. The *post hoc* analyses showed that participants ascribed more mental states for all five dimensions of the AMS after the interaction with Pepper, conversely after the interaction with NAO increased the attribution of the Intention/Desires, Imaginative and Perceptive dimensions, but not of Epistemic and Emotive states (Table 2).

The effect of the interaction on the expectations of the future robots

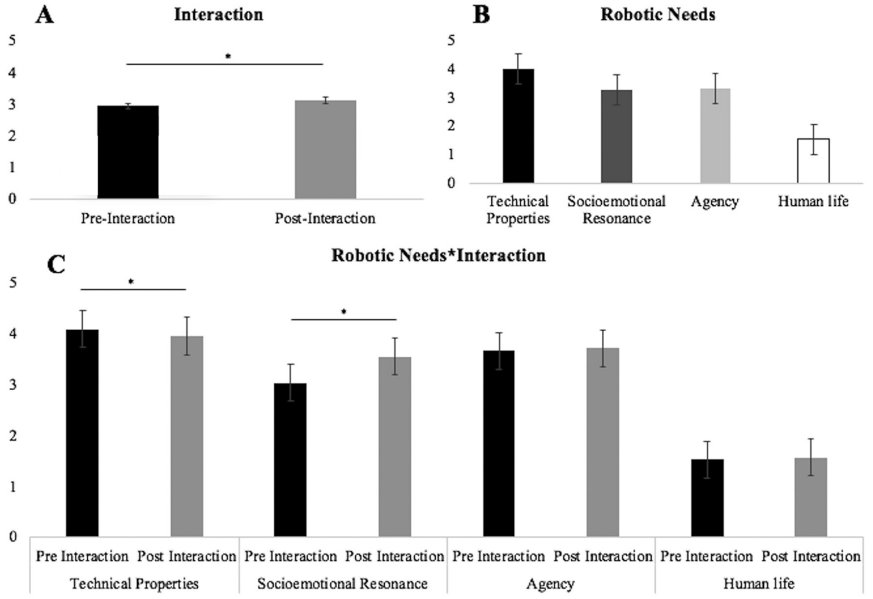
To evaluate the type of robot on people's expectations of future robots, a GLM analyses was carried out with two levels of *Interaction* (pre-interaction, post-interaction) and four levels of *Robotic Needs* (technical advantages, socioemotional resonance, agency, human life) as within-subject factors and *Robot* (NAO, Pepper) as between-subject factor. The GLM revealed (1) a significant main effect of *Interaction* (F=8.74, p<0.005), suggesting that participants' expectations are higher after interaction compared to the expectations before interaction (post-interaction > pre-interaction; MDiff=–0.116, SE=0.039, p<0.005; Figure 3A on page 205); (2) a significant main effect of *Robotic Needs* (F=522.26, p<0.001; Figure 3B), suggest-

Table 2 - *Statistics comparing the attribution of AMS dimensions as a function of the robot (NAO, Pepper) and interactions (pre-interaction, post-interaction)*

	NAO			Pepper		
AMS dimensions	*Mean Diff. pre-interaction vs. post-interaction*	*SE*	*Sign. b*	*Mean Diff. pre-interaction vs. post-interaction*	*SE*	*Sign. b*
Epistemic	0.255	0.208	0.223	–0.473	0.178	0.009
Emotive	–0.354	0.199	0.077	–0.748	0.17	0.001
Desires/intentions	–0.436	0.207	0.037	–0.787	0.178	0.001
Imaginative	–0.727	0.204	0.001	–0.838	0.175	0.001
Perceptive	0.915	0.23	0.001	–0.129	0.197	0.514

ing that the technical advantages are the higher expectations for the development of the new robots, conversely the human life features are the lower expectations for the development of the next generation of robots.

Figure 3 (A-C) - *Bar chart representing: (A) the SRN mean scores as a function of interaction; (B) the SRN mean scores as a function of Robotic Needs; (C) the SRN mean scores as a function of interaction and robotic needs*



Note: *Indicates significant differences. SRN, Scale for Robotic Needs.

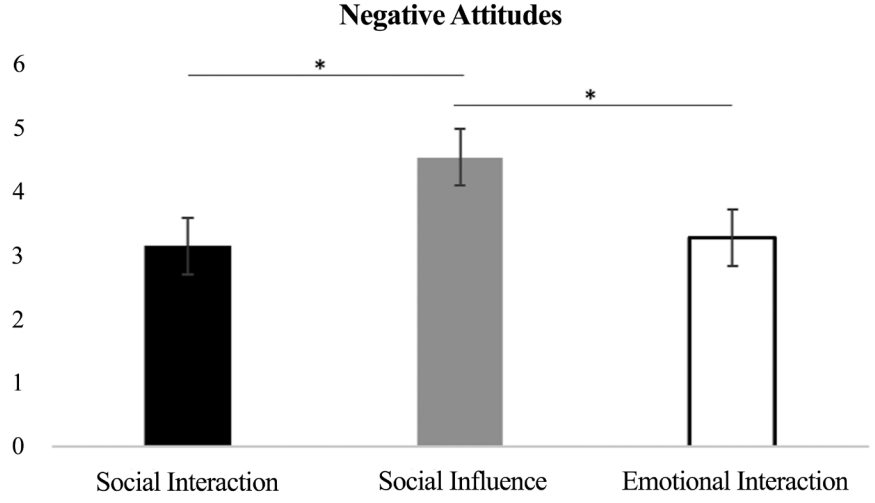
Furthermore, the analysis showed a two-way interaction between *Robotic Needs* and *Interaction* $F(3, 155) = 25.66$, $p < 0.001$, *partial*-$\eta^2 = 0.137$, $\delta = 1$. The *post hoc* analyses showed that participants' expectations are higher for the agency and the socioemotional resonance after the interaction, but the technical advantages decreased after interaction (Figure 3C).

The effect of the interaction on the negative attitudes

To assess the effect of the type of robot people' expectations of future robots a GLM analyses was carried out with two levels of *Interaction* (pre-interaction, post-interaction) and three levels of *Negative Attitudes* (social interaction, social influence, emotions) as within-subject

factors and *Robot* (NAO, Pepper) as between-subject factor. The GLM revealed a significant main effect of *Negative Attitudes*, $F=110.85$, $p<0.001$, *partial*-$\eta^2=0.425$, $\delta=1$, suggesting that the participants are more frightened by the social influence that robots can have on humans compared to both anxiety of social and emotional interactions (Figure 4). However, there are no significant interactions between *Negative Attitudes*, *Interaction*, and *Robot* ($p>0.05$), suggesting that the people's negative attitudes are independent of assisting to an interaction and to the type of robots.

Figure 4 - *Bar chart representing the NARS mean scores (social interaction, social influence, and emotional interaction)*



Notes: *Indicates significant differences. NARS, Negative Attitude toward Robots Scale.

Discussion

The present study investigated the variability of the AMS, expectations, and negative attitudes toward future robots as function of the physical appearance of two HSRs, NAO, and Pepper, after observing a real interaction with a human (an experimenter).

The results of the AMS showed that there were no differences between the two HSRs before the interaction. These findings are in line with a previous study showing that these two HSRs are similar with respect to their level of human likeness (Phillips et al., 2018). However,

the scenario substantially changed after the interaction, revealing the human's sensitivity to the physical appearance of the HSRs. This sensitivity in terms of AMS was already observed in early childhood, highlighting a developmental trend indicating that children at the age of 5 are less sensitive to the physical differences of robots than older children (Manzi et al., 2020b).

The results of the present study prove that this sensitivity to the physical appearance of robots is fully acquired in adulthood. In particular, participants attribute higher mental qualities to all dimensions of the AMS after witnessing the interaction with the Pepper robot, indicating that the design of the Pepper robot evokes higher mental properties than the NAO robot. Although this result seems to be in contrast with the findings of the study by Thunberg et al. (2017), which showed a greater increase in positive attributes to the NAO than to the Pepper, it is important to highlight that the two interactions are substantially different.

The interaction implemented for this study was more friendly with respect to the interaction of the study by Thunberg et al. (2017) which was characterized by the oppositional behavior of robots toward the human partner. Therefore, future studies should address the effects of these two robots in different scenarios.

The results of the negative attitudes revealed that these are independent of both the interaction and the type of HSRs. These data showed, on the one hand, that the direct experience of a single interaction with robots is not sufficient to trigger a change in negative attitudes toward robots, and, on the other hand, that attitudes are not influenced by the subtle design difference of these two HSRs. Therefore, future studies should further investigate the effect of repeated interactions with a wider range of physical differences of the HSRs, from the more mechanical to the more human.

The trends observed in our study regarding the AMS after witnessing an interaction between a human and a robot are also found in relationship to people's expectations of future robots. However, expectations are independent of the type of robot, highlighting that these are associated with a more general view of robotic technologies.

In particular, expectations regarding mechanical properties decrease after interaction, showing how people who experience the real functioning of social robots are more aware of current technological advances. Moreover, people expect robots to improve their social abilities after interaction. This result may plausibly depend on the people's general misconception that includes both robotics and artificial intelligence (AI) in a unique theoretical framework; in other words, people assuming that a robot necessarily incorporates AI expect to improve social skills after in-

teraction. In this sense, further studies are required to grasp this possible overlap between robots and AI.

With respect to robot humanization expectations, these represent the least considered and desired properties for robots of the future. This result seems to confirm the hypothesis that although people desire robots to become more efficient in terms of interactional abilities, at the same time, they are particularly worried about their extreme humanization (Manzi et al., 2021). Overall, the findings regarding people's expectations for the development of future robots are in line with a previous study (Manzi et al., 2021) which shows that people generally prefer the development of technical and interactive components rather than a greater humanization of future robots. Furthermore, this study enriches the topic of expectations on future robots, showing how direct experience with robots can change people's expectations (Inghilleri et al., 2015).

To sum up, this study showed that humans are particularly sensitive to the design of HSRs in terms of attribution of mental qualities even when the HSRs differ slightly in their physical appearance. Furthermore, the results indicated that negative attitudes toward robots are not influenced by either the interaction or the type of HSRs, showing that these attitudes may probably change as function of repeated interactions. Finally, the results revealed that the observation of an interaction influences people's expectations for the development of future robots, in particular increases their awareness of the real technical capabilities of robots and increases their desire for a more interactive robot. In conclusion, although further studies are needed, we can claim that it is important to overcome the universal assumptions about robots and to assume a new perspective that not all robots are equal.

References

Ainsworth, M. D. S. (1969). Object relations, dependency, and attachment: A theoretical review of the infant-mother relationship. *Child development, 40*(4), 969-1025.

Ainsworth, M. D. S., Bell, S., & Stayton D. J. (1971). Individual differences in strange situation behavior of one year olds. In H. R. Schaffer (Ed.), *The Origins of Human Social Relations* (pp. 17-58). Academic Press.

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science robotics, 3*(21).

Beuscher, L. M., Fan, J., Sarkar, N., Dietrich, M. S., Newhouse, P. A., Miller, K. F., & Mion, L. C. (2017). Socially assistive robots: measuring older adults' perceptions. *Journal of gerontological nursing, 43*(12), 35-43.

Bowlby, J. (1969). *Attachment and loss, Vol. 1: Attachment.* Basic Books.

Bowlby, J. (1973). *Attachment and loss, Vol. 2: Separation*. Basic Books.

Bowlby, J. (1980). *Attachment and loss, Vol. 3: Loss, sadness and depression*. Basic Books.

Dario, P., Guglielmelli, E., & Laschi, C. (2001). Humanoids and personal robots: Design and experiments. *Journal of robotic systems*, *18*(12), 673-690.

Di Dio, C., Manzi, F., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., & Marchetti, A. (2020a). It does not matter who you are: Fairness in pre-schoolers interacting with human and robotic partners. *International Journal of Social Robotics*, *12*(5), 1045-1059.

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., & Marchetti, A. (2020b). Come i bambini pensano alla mente del robot: il ruolo dell'attaccamento e della Teoria della Mente nell'attribuzione di stati mentali ad un agente robotico [How children think about the robot's mind. The role of attachment and Theory of Mind in the attribution of mental states to a robotic agent]. *Sistemi Intelligenti*, *1*(20), 41-56.

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., & Marchetti, A. (2020c). Shall I trust you? From child-robot interaction to trusting relationships. *Frontiers in psychology*, *11*, 469.

DiSalvo, C. F., Gemperle, F., Forlizzi, J., & Kiesler, S. (2002, June). 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* (pp. 321-326).

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and autonomous systems*, *42*(3-4), 177-190.

Edwards, A., Edwards, C., Westerman, D., & Spence, P. R. (2019). Initial expectations, interactions, and beyond with social robots. *Computers in Human Behavior, 90*, 308-314.

Inghilleri, P., Riva, G., & Riva, E. (Eds.) (2015). Introduction: Positive change in global world: Creative individuals and complex societies. In P. Inghilleri, G. Riva & E. Riva (Eds.), *Enabling Positive Change Flow and Complexity in Daily Experience* (pp. 1-5). De Gruyter Open.

MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, *7*(3), 297-337.

Manzi, F., Ishikawa, M., Di Dio, C., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., & Marchetti, A. (2020a). The understanding of congruent and incongruent referential gaze in 17-month-old infants: An eye-tracking study comparing human and robot. *Scientific Reports*, *10*, 11918.

Manzi, F., Peretti, G., Di Dio, C., Cangelosi, A., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., & Marchetti, A. (2020b). A robot is not worth another: Exploring children's mental state attribution to different humanoid robots. *Frontiers in Psychology*, *11*, 2011.

Marchetti, A., Di Dio, C., Manzi, F., & Massro, D. (2020). Robotics in clinical and developmental psychology. *Reference module in neuroscience and biobehavioral psychology*, B978-0-12-818697-8.00005-4.

Marchetti, A., Di Dio, C., Massaro, D., & Manzi, F. (2020). The psychosocial fuzziness of fear in the coronavirus (COVID-19) era and the role of robots. *Frontiers in Psychology*, *11*, 2245.

Marchetti, A., Manzi, F., Itakura, S., & Massaro, D. (2018). Theory of mind and humanoid robots from a lifespan perspective. *Zeitschrift für Psychologie, 226*(2), 98-109.

Mori, M. (1970). The Uncanny Valley. *Energy*, 7, 33-35.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, *19*(2), 98-100.

Nomura, T., Suzuki, T., Kanda, T., & Kato, K. (2006). Measurement of negative attitudes toward robots. *Interaction Studies, 7*(3), 437-454.

Pandey, A. K., & Gelin, R. (2018). A mass-produced sociable humanoid robot: Pepper: The first machine of its kind. *IEEE Robotics & Automation Magazine, 25*(3), 40-48.

Phillips, E., Zhao, X., Ullman, D., & Malle, B. F. (2018). What is human-like? Decomposing robots' human-like appearance using the anthropomorphic robot (abot) database. *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*, 105-113.

Prakash, A., & Rogers, W. A. (2015). Why some humanoid faces are perceived more positively than others: Effects of human-likeness and task. *International journal of social robotics, 7*(2), 309-331.

Robins, B., Dautenhahn, K., & Dubowski, J. (2006). Does appearance matter in the interaction of children with autism with a humanoid robot?. *Interaction studies, 7*(3), 479-512.

Sinnema, L., & Alimardani, M. (2019). The attitude of elderly and young adults towards a humanoid robot as a facilitator for social interaction. In M. A. Salichs, S. S. Ge, E. I. Barakova et al. (Eds.), *Social Robotics. Vol. 11876. Lecture Notes in Computer Science* (pp. 24-33). Springer.

Thunberg, S., Thellman, S., & Ziemke, T. (2017, October). Don't judge a book by its cover: A study of the social acceptance of NAO vs. Pepper. In *Proceedings of the 5th International Conference on Human Agent Interaction* (pp. 443-446).

Wu, Y. H., Fassert, C., & Rigaud, A. S. (2012). Designing robots for the elderly: Appearance issue and beyond. *Archives of gerontology and geriatrics*, *54*(1), 121-126.

Zanatto, D., Patacchiola, M., Cangelosi, A., & Goslin, J. (2020). Generalisation of anthropomorphic stereotype. *International Journal of Social Robotics, 12*(1), 163-172.

SECTION 3

Field Research in Human-Robot Interaction

1. Artificial Intelligence and AI-guided Robotics for Personalized Precision Medicine

V. Valentini, M. D'Oria, A. Cesario

ABSTRACT

Modern medicine is pursuing a Personalized Precision Medicine approach, aimed at providing 'the right treatment, to the right patient, at the right time', to analyze a heterogeneous ecosystem of variables concerning the disease and the person. Artificial Intelligence (AI) and AI-guided Robotics play a promising role in diagnostics and therapeutics. Despite most current innovative solutions not yet being advanced enough for clinical use, research is focused on using machine learning algorithms for several purposes. Considering the clinical scenario of the Fondazione Policlinico Universitario A. Gemelli IRCCS and, in particular, the real case of the 'Gemelli Generator' project, the aim of this chapter is to show four possible applications of AI and AI-guided robotics available for research and/or clinical practice. Two forthcoming scenarios (in silico clinical trials and digital therapeutics) in Personalized Precision Medicine are described, and some of the main challenges facing the use of AI are highlighted.

Introduction

The health ecosystem is dynamic and reflects the conditions, both internal and environmental, to which a person is exposed. Several human diseases are complex and multifactorial; therefore, they should be understood at a systemic as well as molecular level (Jain, 2009). The completion of the Human Genome Project (National Human Genome Research Institute, 2003), through which the entire human genome was mapped, opened the doors to a new understanding of human diseases by providing precious information on person-specific PATO phenotypes.

Since then, medicine has started pursuing a Personalized Precision Medicine approach, aimed at providing 'the right treatment, to the right patient, at the right time': to achieve this goal, the entire ecosystem of variables that may affect a patient's health needs to be considered. The identification of precise 'omic' biomarkers (e.g., genomics, proteomics, metabolomics) with quantitative imaging analysis and radi-

omics applications, gained much interest in the scientific community by enabling the development of data-driven evidence to support decision-making (Alyass et al., 2015; Valentini & Cesario, in press). On the other hand, the retrieval of heterogeneous digital data (Real-World Data, RWD) from several sources (e.g., electronic health records, wearable devices, medical apps - Internet-of-Medical-Things, IoMT) has deepened the understanding of diseases on both an individual and population level. Massive digital data are known as 'Big Data' and respect the following characteristics (the '5 V rule') (Valentini & Cesario, in press):

– *Volume*: data must be numerous.
– *Variety*: data must be heterogeneous.
– *Veracity*: data and their sources must be reliable (e.g., imaging, electronic health record, lab, and microbiological data).
– *Value*: data must be important and relevant for the goal of the study.
– *Velocity*: data must be retrieved, analyzed and rapidly accessible.

Both instances ('omics' and RWD) help provide a compelling picture of the person in order to find the most appropriate therapy or preventive solution. In this scenario, Artificial Intelligence (AI) is increasingly used to manage information and encourage evidence generation from RWD (also known as Real-World Evidence, RWE) (Marazzi et al., 2021). AI solutions (defined as specific machine learning [ML] and deep learning [DL] algorithms) can help in analyzing and managing great quantities of patients' clinical and digital information, to predict the etiology of diseases as well as prevent the outcome of certain therapies (Boldrini et al., 2019). As reported in Table 1, several types of algorithms are used to study different things.

Table 1 - *Types of algorithms with practical examples from the literature in diagnostics and therapeutics*

Type	*Practical examples of ML usage in diagnostics*
Linear regression	It is applied in numerous analysis and computational predictions, from the identification of relevant prognostic risk factors (Schneider et al., 2010) to the understanding of prevalent patterns in human HIV immunodeficiency (Madigan et al., 2008).
Logistic regression	It is applied to obtain predictive and explicative models to generate risk profiles in complex diseases or cardiac pathologies (Xu et al., 2018).
Discriminant analysis	It is applied for the early detection of Parkinson's disease symptoms (Armañanzas et al., 2013) or to evaluate the risk for chronic diseases (Jen et al., 2012).

Support Vector Machine (SVM)	It is applied in several scenarios, from symptom classification for diagnostic accuracy to the identification of biomarkers through tumor imaging, genetic and genomic profiling (Huang et al., 2018; Cho et al., 2019; Cruz & Wishart, 2007; Huang et al., 2018; Ahmed et al., 2020).
Naïve Bayes	It is applied to model several tumor types [brain, prostate, breast (Langarizadeh & Moghbeli, 2016) and other diseases [e.g., Alzheimer's disease (Wei et al., 2011)] and to support clinical decision-making in cardiac diseases (Srinivas et al., 2010).
K-nearest neighbor (K-NN)	It is applied to model diagnostic performance (Zhang, 2016), to classify health records (Al-Aidaroos et al., 2012) or to predict pancreatic cancer (Zhao & Weng, 2011).
Deep Learning	It is applied to diagnostics in several clinical fields, especially oncology (Hosny et al., 2014; Langlotz et al., 2019), dermatology (Olsen et al., 2018), and cardiology (Rajkomar et al., 2018), by analyzing data from radiography and genomic biomarkers.
Hidden Markov Model (HMM)	It is applied to model a patient's status (Wall & Li, 2009) or to minimize adverse events to drugs (Huang et al., 2018) as well as to monitor circadian activity (Huang et al., 2018).
Decision tree	It is applied to support clinical decision-making and diagnostics (Bae, 2014), as well as to understand factors that cause hypertension (Tayefi et al., 2017) or pressure ulcers in elderly patients (Moon & Lee, 2017).
Random forest	It has several applications, including Alzheimer's disease diagnosis (O'Bryant et al., 2010) and classification (Byeon, 2019), predicting patient mortality in Intensive Care Units (Lee, 2017) or metabolic pathways (Toubiana et al., 2019).
Genetic algorithm	It is applied in radiology, oncology, cardiology, endocrinology, pediatrics, surgery, pulmonology, gynecology, orthopedics, and other important fields (Ghaheri et al., 2015) to predict outcomes and develop noninvasive diagnostic techniques.

Predictive models are based on RWD which are classified in order to attain analogies, rules, connections, neural networks, statistics, or probabilities. According to the literature, the main tasks assigned to algorithms for diagnostics and decision-making are the following:

– using ontologies (shared and detailed models of a problem or domain) to create semantic models of data by connecting basilar instances (Zouri et al., 2019);

– risk prediction and diagnosis of several diseases (especially onco-

logical) according to their typologies, features, and levels of complexity (Kourou et al., 2014);

– classification of cancer types towards visual evaluation of tumoral cells (Sun et al., 2018; Koelzer et al., 2019);

– multi-Omics Data Integration to personalize the diagnosis of complex diseases (Patel et al., 2020; Cesario et al., 2021a);

– clustering of biomarkers as potential PATO phenotypes (Hedman et al., 2020; Sun et al., 2020);

– identification of new casual and causative disease pathways to generate new hypotheses (Naylor, 2018);

– data clustering of biological pathways for pathway modeling and analysis (Zhu et al., 2017; Zhang et al., 2017);

– identification of molecular and genomic elements which may be sensitive to existing or innovative treatments (Fornecker et al., 2019) to predict adverse events;

– identification of new associations between diseases and their causes, including comorbidity and multimorbidity scenarios (Zamborlini et al., 2017; Deparis et al., 2018);

– identification of non-linear relationships in the electronic health record (e.g., towards text mining) (Brunekreef et al., 2021; van Dijk et al., 2021);

– visualization and analysis of knowledge structures extracted from the IoMT to identify explicit and hidden patterns (Sangaiah et al., 2020);

– observation of disease-specific therapeutic outcomes, to identify new instances (e.g., mutations, genetic alterations) that cause new diseases (Fornecker et al., 2019);

– enhancement of screening to avoid surgery or other invasive treatments (Franceschini et al., 2021).

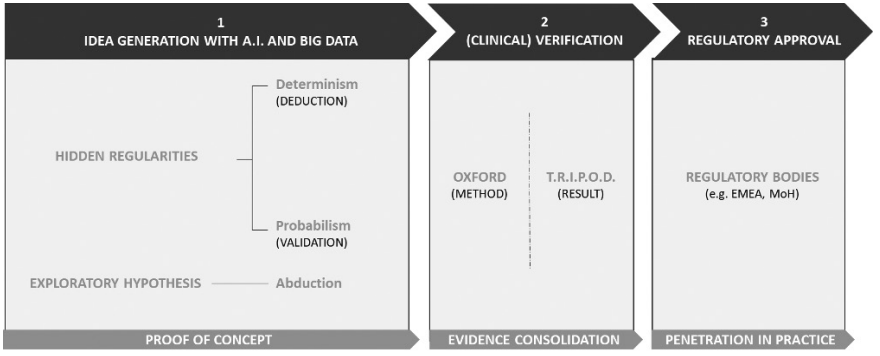
The medical areas in which more literature is available are radiology, radiotherapy, ophthalmology and dermatology (Naylor, 2018), but also gastroenterology, gynecologic oncology and breast cancer oncology (Sangaia et al., 2020), hematology (Radakovich et al., 2020), and infectious diseases (Peiffer-Smadja et al., 2020) including COVID-19 (Ozsahin et al., 2020). The process for AI penetration within clinical practice is not straightforward, but can be simplified as follows (Figure 1):

1. Step 1. Most research is focused on idea generation by using AI and Big Data to find hidden regularities among data (aiming for a deductive or probabilistic approach) or to explore hypotheses (aiming for an abductive approach). The outcome of this step is to generate a proof of concept (PoC).
2. Step 2. Available PoCs can undergo a clinical verification process, which may follow the Oxford approach (whose aim is to validate

the method used by the PoC) or the T.R.I.P.O.D. approach (whose aim is to validate the result of the PoC). Both approaches lead to evidence consolidation.

3. Step 3. Very few cases of consolidated evidence have undergone a regulatory process by regulatory bodies (e.g., EMA, AIFA, etc.), which would allow the penetration of the AI-solution into clinical practice as the ultimate outcome.

Figure 1 - *Steps to a validated and regulated application of AI-based technologies within clinical practice*



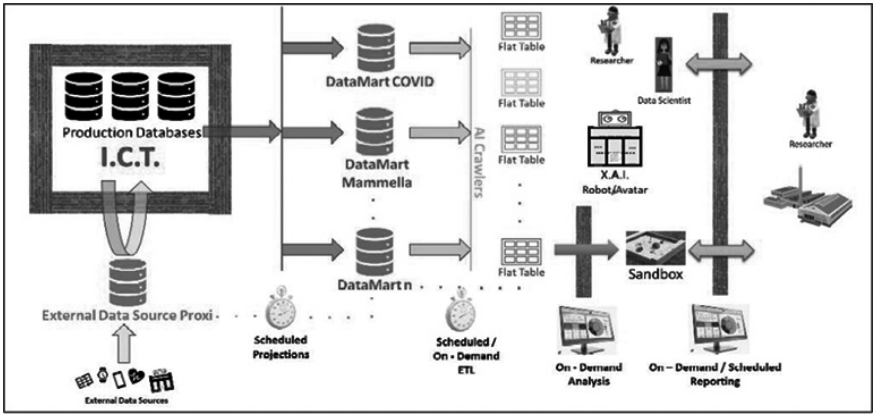
The Real Case Scenario of Gemelli Generator

In a world enriched by data available from different sources (e.g., apps that monitor one's health status, wearable devices, etc.), every healthcare organization must collect, store and analyze large volumes of patient information, to answer different challenges of care among populations, while finding adequate clinical solutions. It is crucial to possess advanced informatics technologies to understand and use RWD to generate RWE: these technologies may help clinicians and researchers process data in a more precise and reliable way.

To achieve this goal, since 2019 the research hospital, Fondazione Policlinico Universitario A. Gemelli IRCCS (FPG), has implemented a specific research infrastructure (called 'Gemelli Generator') to meet research and care needs through dedicated services that improve clinical practice. The goal of Gemelli Generator is to clinically and scientifically valorize skills, processes and data acquired over the years within the Data Warehouse (DWH) and in FPG's biobank, through advanced services of data collection, extraction and analysis. The DWH has collected over

760 million data since 1999: figure 2a shows the process of data management for researchers' use, from the database to the research product (e.g., data analysis, predictive model, customized bot, avatar, etc.); figure 2b shows the main sources of these data.

Figure 2a - *The architecture of Gemelli Generator*



Note: The blockhouse represents the DWH of FPG, monitored by the Information and Communication Technology Unit. The DWH can be populated by data from FPG (see Figure 2b) as well as external data sources. Data can be managed and clustered in several data marts created for specific contents (e.g., COVID, breast cancer, etc.). AI crawlers re-arrange data into flat tables, according to the necessities of the researcher. The outcome of a flat table can a) be shared with the researcher, b) be managed by a data scientist at FPG to share results with external researchers, c) generate a virtual avatar, d) populate a sandbox where external researchers and industries can request on-demand analysis or reports. The whole architecture is compliant with current General Data Protection Regulation (GDPR), safe-by-design, and fully integrated with the hospital Information and Communication Technology (ICT) infrastructure.

The mission of Gemelli Generator is to support several healthcare assets within a strategic interdisciplinary model, from basic to clinical research, by focusing on each patient's care outcomes. The project is developed along the lines of four research core facilities:

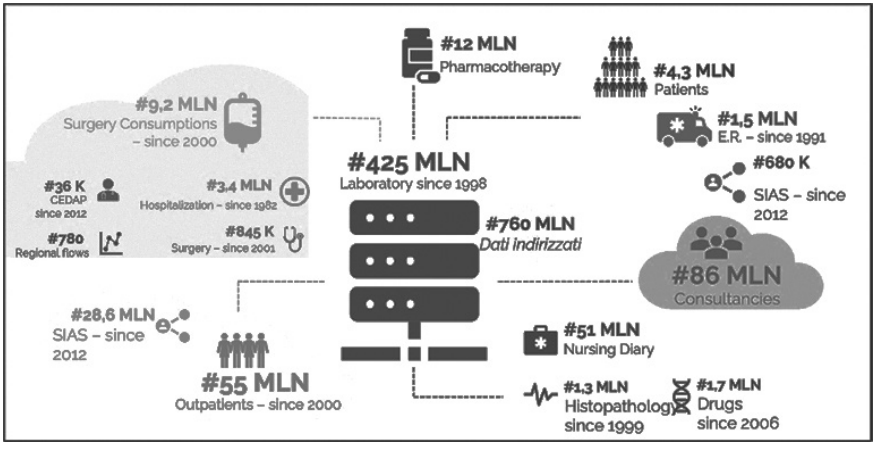
– Bioinformatics: this offers services for the implementation and development of workflows suitable for personalized bioinformatics analysis in every research activity, starting from raw or pre-processed data through to new and experimental data, as well as data retrieved from public databases.

– Data Collection: this offers data collection services as well as methodological, educational, analytic, and data management support according to Good Clinical Practice (GCP), the GDPR, and accuracy, completeness, consistency, integrity and timing (ACCIT) for data quality.

– Epidemiology & Biostatistics*:* this offers methodological counselling for correct research planning and development, statistical analysis and data management.

– Real World Data: this offers Big Data extraction, analysis and processing services for creating enhanced AI/ML techniques, in order to develop automated systems and predictive models that can support decision-making.

Figure 2b - *An overview of FPG sources that help in populating its DWH*



Given the focus of this chapter, the services provided by the Real-World Data facility are shown in Table 2.

Several achievements can be reached in the prediction, prevention, precision and personalization of care, as well as in patients' and caregivers' participation in the care journey. Researchers can use numerous data and information otherwise not accessible, but also with professional support at all phases of the research, to improve clinical practice while increasing knowledge. Patients can receive more efficient care in terms of clinical processes as well as personalized treatments, both characterized by patient-specific data to maximize the results.

In the following paragraphs, we will consider some AI and robotics applications in clinical and research practice, even though the scenario of AI implementation and usage is wider than that:

a) AI-robot assisted surgery and radiotherapy for mini-invasive treatments.

b) ML/DL algorithms for outcome prediction, assessment, and patient enrollment in clinical trials.

c) Process mining algorithms for pathways and clinical trials appropriateness.

d) Patient Support Programs for patient journey continuity.

Table 2 - *Main services of the Real-World Data facility*

Service	*Description*	*Research field*
Data Clustering	Data mart organized by pathology from different sources (labs, radiomics, omics) based on data/text mining	Observational clinical trials Network projects (multicentric) Data products for academic and industrial collaborations
Predictive Models	BOT: automated ML system and data analysis and visualization AVATAR: patient's virtual model for clinical pathways simulation	Early diagnostics: pathology risk and complication ('alert') Clinical pathway optimization to improve care and resources Prognostic support: simulation of treatment response
Support to Clinical Trials	Data management systems, ML, and telemedicine to support interventional clinical trials	Smart patients' enrollment Simulation of treatment response Integrated e-care systems

a) AI-robot assisted surgery and radiotherapy for mini-invasive treatments

Mini-invasive surgery (laparoscopy, robotic surgery) helps to reduce post-operative pain and recovery time after intervention (Ashrafian et al., 2017). In 2018, FPG activated the Gemelli Robotics Mentoring Center (GeRoMe) with robotic surgery as its main goal. Surgical theaters were equipped with latest generation surgical robots and one of the most modern systems of electrosurgery, which includes the use of radiofrequency and ultrasound to help surgeons conduct mini-invasive laparo-

scopic, robotic, and endoscopic interventions. Integrated surgical theaters also have 3D/ICG and 4K vision systems, displayed on 55° 43/SD monitors that can be managed remotely by way of touch screens, with pre-uploaded scenarios and videoflows in full 4K Over IP.

AI-robot-assisted surgery helps the surgeon perform precision procedures and, despite most AI-technologies still enjoying little use in clinical practice (due to the regulation process and the reliability/readiness of tools), some AI-robot-assisted systems are still available at FPG for minimally invasive treatments (see Table 3).

Table 3 - *Description of two AI-robot assisted technologies for minimally invasive treatments at FPG*

Technology	*Description*
The Da Vinci ® Surgical System (Intuitive Surgical Inc.)	A 3D vision system that allows the capture of accurate images. Robotic arms can replicate human arms through an advanced motion control system in order to perform complicated surgeries, performing small incisions that mean less pain for patients and, therefore, faster recovery. The ergonomic design also offers the surgeon more comfort so as to reduce fatigue-related problems (Yi et al., 2021).
MRIdian ® (ViewRay Inc.)	A hybrid linear accelerator that acquires MRI images during radiation, while monitoring the position of the tumor and surrounding healthy tissues in real-time. Integrated MRI enables a better visualization and synchronizes radiation beaming with the physiological movements of the organ (e.g., respiration, heartbeat, visceral movements); in this way, the radiation beam hits the disease precisely without touching healthy tissues, so reducing collateral effects and maximizing therapeutic outcomes.

b) ML/DL algorithms for outcome prediction, assessment, and patient enrollment in clinical trials

While retrieving patients' data from wearable devices, diagnostic tests, and other relevant sources, the hospital database collects and stores information with a strong potential for clinical research goals. Hence, a researcher may be interested in understanding how stratified genomic variables on a disease are related to demographics and lifestyle data in the disease expression. ML/DL algorithms could give access to the heterogeneity of data from 'similar' patients and model disease progression.

ML/DL algorithms help to acquire, analyze, predict and assess clinical trial outcomes (Cesario et al., 2021b), so as to prevent and prematurely manage potential adverse events (e.g., toxicity, aggravation, relapse), en-

suring the best patient journey based on the evidence and therefore personalizing the treatment. Some algorithms can mine information from the hospital database by extracting them according to specific variables indicated by the researcher/clinician and populate a customized data mart. Other algorithms will crawl into the data mart to select information and generate flat tables (such as excel documents), in order that specific bots provide the researcher/clinician with the required information. Table 4 shows three examples of current studies running at FPG that use ML/DL algorithms for outcome prediction and assessment in clinical trials.

Table 4 - *List of three current studies using AI-solutions for outcome prediction and assessment in clinical trials at FPG*

Study	*Pathology*	*Description*
GENERATOR Tracer RT PI: Prof. V. Valentini Radiation Oncology	Cancer	A clinical trial using AI-supported outcome prediction and assessment monitoring technologies to estimate the risk of healthcare professionals and cancer patients being infected by the SARS-COV-2 virus, while safely managing their needs.
AI-APACHE PI: Prof. G. Scambia Gynecologic Oncology	Gynecologic cancer	A clinical trial using AI-supported outcome prediction and assessment monitoring technologies during multimodal oncological therapies and follow-up periods of cervical cancer patients.
GENERATOR INTERFACE PI: Prof. A. Cingolani Infectious Diseases	HIV	A clinical trial using AI-supported outcome prediction and assessment monitoring technologies for patients with HIV.

Clinical researchers can also be helped in recruiting patients for clinical trials through a specific platform (called 'Digital Research Assistant', DRA), which is supported by an algorithm that performs a matchmaking between patient's minimal characteristics (e.g., BRCA1/2 mutations, age, molecular characteristics, clinical stage, etc.) and a clinical trial's inclusion criteria so as to personalize patient enrollment (Cesario et al., 2021b).

c) Process discovery algorithms for pathways and clinical trial appropriateness

Healthcare delivery can be a complex task in high-volume hospitals. Analysis of Big Data using process discovery (or process mining) algo-

rithms can help analyze different information (e.g., demographics, imaging data, field notes, laboratory results) to create predictive models for disease risk, diagnosis, treatment and treatment appropriateness. Process mining algorithms are crucial for describing, and therefore optimizing, patient flows for care delivery (Ngyam & Khor, 2019), for clinical trial management, and for guideline conformance checking. Table 5 shows some studies running at FPG that implement process mining algorithms for pathways and clinical trial appropriateness.

Table 5 - *List of three current studies using process mining algorithms running at FPG*

Study	*Short description*
Palliative patient flow in radiotherapy department	500 palliative radiation treatment plans of patients with bone and brain metastases were included in the study, corresponding to 290 patients treated in our department in 2018. Event-logs and the relative attributes were extracted and organized. A process discovery algorithm was applied to describe the real process model, which produced the event-log. Finally, conformance checking was performed to analyze how the acquired event-log database works in a predefined theoretical process model (Placidi et al., 2021).
First prototype of the 'Multidisciplinary Tumor Board Smart Virtual Assistant'	The prototype aimed to achieve 1) Automated classification of clinical stage starting from different free-text diagnostic reports; 2) Resolution of inconsistencies by identifying controversial cases drawing the clinician's attention to particular cases worthy for multi-disciplinary discussion; 3) Support an environment for education and knowledge transfer to junior staff; 4) Integrated data-driven decision-making and standardized language and interpretation (Macchia et al., submitted).
Description of a methodology to deal with conformance checking through the implementation of computer-interpretable-clinical guidelines (CIGs)	The purpose of this paper was to describe a methodology to deal with conformance checking through the implementation of computer-interpretable-clinical guidelines (CIGs) and present an application of the methodology to real-world data and a clinical pathway for radiotherapy-related oncological treatment. Three usage cases were presented, in which the results of conformance checking were used to compare different branches of the executed guidelines with respect to the adherence to ideal process, temporal distribution of state-to-state transitions, and overall treatment (Lenkowicz et al., 2018).

d) Patient Support Programs for patient journey continuity

Another important use for AI in both clinical and research practices would be in the implementation of a more complex and integrated assistance to patients, through the creation of 'virtual hospitals' to ensure care continuity. Patient Support Programs (PSPs) are digital solutions aiming at ameliorating patient management by offering dedicated pathways for those who risk relapse or a deterioration in their health, so threatening the efficacy of treatments. These patients may have several unmet needs during their journey, such as:

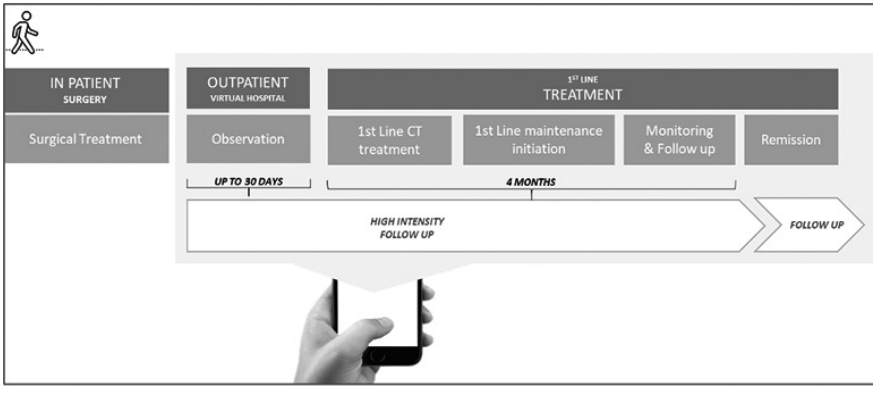
– interacting with different health professionals who do not work in a team, leaving patients with a sense of service fragmentation;

– high-expenditure of money and time for first-line treatments and in-person consultations;

– a discontinuous monitoring, with the risk of not anticipating potential adverse events or verifying actual treatment adherence.

To attenuate these issues, digital tools like apps and platforms are integrated into the care program to monitor patients' health while offering a professional sustain during the patient journey (Figure 3).

Figure 3 - *The image represents an example of a PSP intervention for an oncological patient, in order to guarantee continuous assistance and monitoring even when the patient is not physically at the hospital*



PSPs should not be confused with classical telemedicine, as they amplify the concept of assistance by creating a 24/7 support for the patient with a network of healthcare and non-healthcare professionals, as well as caregivers, so providing the best assistance through solutions such as:

– a patient-centered digital platform;
– data centralization for real-time availability;
– dedicated care managers in a digital environment;
– televisits and teleconsultations to reduce time and costs, real-time telemonitoring to verify treatment adherence and ensure follow-up;
– AI/ML systems to analyze data and create predictive models to monitor the patient's health.

The patient can have a point of professional reference fully accessible for clinical support, which can impact on his/her quality of life (QoL). Specifically, PSPs can offer several benefits like:

– the involvement of several professionals to identify the contents and modalities of personalized patient assistance;
– the management of therapy in an optimal way, to improve patient's QoL;
– continuity of follow-up, letting the patient feel in charge even when at home;
– a reduction in hospital visits and travel;
– answers to the patient's doubts and worries by way of virtual coaching and training;
– in-depth and personalized knowledge of the patient's clinical history;
– prevention and reduction of side effects;
– self-management of the disease by adopting a correct lifestyle;
– receiving alerts for drugs about to run out;
– caregiver empowerment and education for patient management;
– at home nurse/health professional to provide training on the use of the device;
– free counselling service for psychological and emotional support;
– easy reservation and management of diagnostic examinations and therapies.

In the context of FPG, for example, the possibility of providing a virtual hospital requires:

– a responsible professional and social activity, because different health systems would have to be integrated and collaborate synergically;
– maximization of therapeutic pathways with a reduction in costs and time;
– patient and caregiver empowerment and engagement, to improve interaction with the hospital;
– valorization of the doctor-patient relationship with global care continuity;
– integrated management of the patient's therapeutic pathway;
– aggregated real-life data that can be used to understand the disease and personalize treatments.

Table 6 shows two ongoing studies at FPG. In both studies, data from the DWH can be used to offer patients the best suited clinical trial for their specific situation.

Table 6 - *Two examples of clinical trials being run at FPG that use AI solutions to monitor patients while retrieving necessary information to provide a patient-centric journey*

Study	*Description*
Virtual management of the follow-up of patients with ovarian cancer and optimization of enrollment in clinical trials	The study uses a patient-centric data capture system to monitor some biometrical parameters of patients with ovarian cancer, with the aim of predicting and preventing risk, as well as allowing patients to signal new symptoms and indicate new information useful for the therapy.
Monitoring of patients with ischemic cardiopathy and heart failure	The study aims to unify technological and environmental aspects, clinical necessities and patient needs in a complex and multidimensional way, towards the creation of eHealth/data capture patient-centric tools that can model individual patient pathways and predict final outcomes so as to personalize therapies and improve patients' QoL.

Forthcoming Scenarios for AI-assisted Personalized Precision Medicine

In the landscape of Personalized Precision Medicine, we envision at least two possible game-changer scenarios that will implement AI solutions to assist in research and care. The first one affects evidence generation by implementing a new approach to clinical research in the form of in silico clinical trials and Human Digital Twins, while the second scenario regards the advent of Digital Therapeutics still considered as medical devices that use AI algorithms as excipients.

Human Digital Twins and in silico clinical trials

Predictive models are already widely and successfully used to reduce the cost and duration of preclinical phases in the development of new bio-

medical products, and current simulation technologies could provide a great benefit to biomedical research in the prediction of disease risk and progression. Consequently, the future scenario of clinical research envisions the introduction of Human Digital Twins (HDTs) for predictive in silico clinical trials (ISCTs).

HDTs are a set of models, which include measurements of a given patient as input parameters and can predict the clinical outcome of medical treatment in that patient. ISCTs are reliable computer models that can be used to study the effect of virtual treatments on virtual patients and predict the product clinical performance in terms of safety and efficacy. The contribution ISCTs will offer in reducing, refining and partially replacing real clinical trials, includes:

– they will help identify different complication risk levels for patients based on their physiological characteristics or provide early evidence of expected behavior of a device. This information could be used to reduce and shorten clinical trials;

– they will enrich the available information on potential outcomes and provide tools for interpreting the possible adverse effects that might be detected;

– clearer correlations between patient characteristics and device interaction could be established;

– they could overcome clinical trial constraints of time and cohort size and predict long term outcomes;

– under the proper circumstances, when the evidence obtained is robust enough, they will partially replace routine clinical trials;

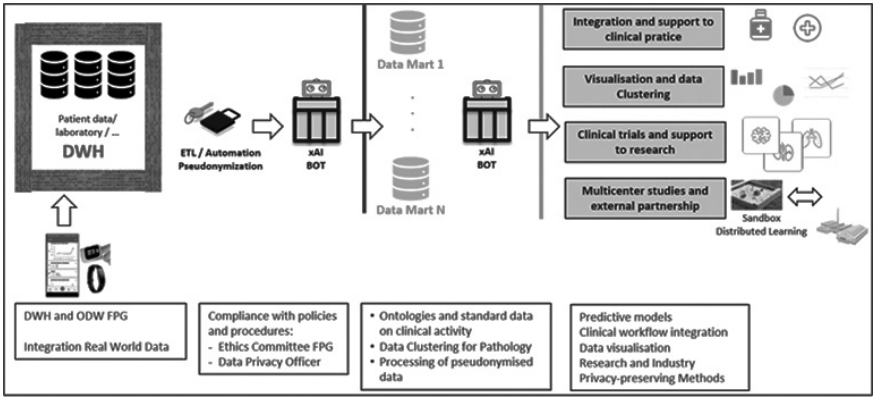
– they will simulate testing conditions difficult to be observed in an actual clinical trial, such as rare effects in very specific populations.

ISCT to test new medical devices will require libraries of patient specific and validated HDTs and computational workflows, which provide a reliable prediction of their clinical behavior in different segments of the target population. Gemelli Generator RWD facility will provide AI-integration of clinical data and the development of predictive models to support diagnosis, treatments, and processes (Figure 4).

An example of the project is the elaboration of a digital avatar implemented to identify potential demographic, clinical, laboratory, radiological and intraoperative variables that can predict complications after colorectal surgery, so creating a predictive score towards a text mining algorithm. Clinicians will access a descriptive dashboard to explore the database and clinical history of about 3000 patients (1500 colon, 1500 rectum), while a digital representation of a patient before surgery will be created to compare the person with classes of similar patients. Hence, predictive models based on the development of a virtual patient (HDT) will estimate length of stay and potential risk of complications. The ex-

pected benefit is greater closeness during patient-doctor interactions in a personalized care philosophy: when needed, patient-specific data will be clearly and cleanly presented, allowing the clinician to focus on the patient rather than data retrieval. Clinicians will also be empowered by letting them retain full control over all available variables while at the same time supporting their decisions with explainable models and clustered retrospective statistics.

Figure 4 - *The image shows how the architecture of Gemelli Generator can dialogue with wearable devices (below) to receive patients' data while anonymization starts the process of the data mart population on specific ontologies (e.g., a disease or another parameter)*



Note: these results can be integrated in clinical practice, visualized and clustered for predictive models, integrated in clinical trials for research purposes, as well as being safely shared with partners involved in multicenter studies

The advent of Digital Therapeutics

Digital Therapeutics (DTx) is a new therapeutic approach intended to encourage a patient's adoption of health-promoting behavior through real-life approaches based on cellphone or computer apps that produce IoMT and, ultimately, to treat diseases (Natanson, 2017). They should not be confused with healthcare and wellness apps available to citizens and patients, because DTx follow specific requirements such as:

– they are recognized as medical devices with a therapeutic goal;
– they are developed through randomized-controlled trials (RCTs);
– they are approved and authorized by regulatory bodies, prescribed by physicians, and reimbursed by healthcare systems;
– they use algorithms as their active principal and excipients.

DTx can be used independently or integrated with other evidence-based treatments (e.g., pharmacological), and their intervention is guided by high-quality software based on scientific evidence, which must be obtained through rigorous and confirmatory clinical trials to prevent, manage, and treat several clinical, mental, and behavioral conditions (Gussoni, 2021). Preliminary clinical applications of DTx are still available and are expected to be more widely implemented in the future, possibly changing the paradigms of treatment (Cho & Lee, 2019).

Main Challenges of the Algorethics Framework

One of the main challenges facing algorethics (the ethics of algorithms) is the use of AI in medicine to analyze large volumes of datasets, which are crucial to training algorithms: the interaction between ML/DL algorithms and Big Data leads to several global topics and reflections. For example, it is important that the reliability and security of data be guaranteed because errors may generate serious consequences for a patient's diagnosis and therapy. Potential ethical and legal implications of directly involving AI in patient care (Franceschini et al., 2021) remain an issue: in fact, a preoccupying concern regards the elaboration of AI decisions in hidden layers, meaning that the logic behind them cannot be verified by a human observer (the so called 'black box problem'). Several questions arise: how much can we delegate decisions to AI? Who should be responsible for its errors? The risk of unintentional outcomes exists, especially when too much trust is placed in an algorithm without supervision and vigilance (Srinivas et al., 2010; Naylor, 2018).

Other challenges concern validation, regulation and assessment issues: the transparency of the algorithm validation and training chain (for example, to reduce 'underfitting' or 'overfitting' biases of predictive models') (Blacklaws, 2018), as well as the creation of a reliable ecosystem that enables industrial, academic, political and social collaboration between the biomedical field and other stakeholders so as to harmonize technical standards that must meet realistic and sustainable goals. With regard to ISCTs and DTx, the Medical Device Regulation (MDR) (EU 2017/745), passed in 2017 and fully applicable in 2021, is expected to see an impressive optimization of clinical trials, by predicting and preventing potential failures, reducing the number of actual patients involved and shortening the duration of follow-up without compromising the study's statistical significance and scientific quality.

Finally, the literacy challenge: it is crucial to establish a process of social awareness, starting from a knowledge and understanding of these AI tools among developers, users (e.g., patients, caregivers) and those who

make decisions (e.g., clinicians, surgeons, nurses) (Schruben, 1980; Valentini & Cesario, in press). To this end, it is necessary to create guidelines for the integration and correct use of AI in diagnostics (Sangaiah et al., 2020), with a proper training of professionals and patients (Steiner et al., 2018) in understanding the meaning of predictive models, without neglecting studies on human-machine interaction and acceptance (Human-Computer Interaction – HCI, User Experience – UX, Technology Acceptance, TA) (FDA, 2019; Paterson & Witzleb, 2018).

References

Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database*, 2020.

Al-Aidaroos, K. M., Bakar, A. A., & Othman, Z. (2012). Medical data classification with Naive Bayes approach. *Information Technology Journal, 11*(9), 1166.

Alyass, A., Turcotte, M., & Meyre, D. (2015). From big data analysis to personalized medicine for all: Challenges and opportunities. *BMC medical genomics, 8*(1), 1-12.

Armañanzas, R., Bielza, C., Chaudhuri, K. R., Martinez-Martin, P., & Larrañaga, P. (2013). Unveiling relevant non-motor Parkinson's disease severity symptoms using a machine learning approach. *Artificial intelligence in medicine, 58*(3), 195-202.

Ashrafian, H., Clancy, O., Grover, V., & Darzi, A. (2017). The evolution of robotic surgery: Surgical and anaesthetic aspects. *British Journal of Anaesthesia, 119*(1), i72-i84.

Bae, J. M. (2014). The clinical decision analysis using decision tree. *Epidemiology and health*, 36, e2014025.

Blacklaws, C. (2018). Algorithms: Transparency and accountability. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376*(2128), 20170351.

Boldrini, L., Bibault, J. E., Masciocchi, C., Shen, Y., & Bittner, M. I. (2019). Deep learning: A review for the radiation oncologist. *Frontiers in oncology, 9*, 977.

Brunekreef, T. E., Otten, H. G., van den Bosch, S. C., Hoefer, I. E., van Laar, J. M., Limper, M., & Haitjema, S. (2021). Text mining of electronic health records can accurately identify and characterize patients with Systemic Lupus Erythematosus. *ACR Open Rheumatology, 3*(2), 65-71.

Byeon, H. (2019). Developing a random forest classifier for predicting the depression and managing the health of caregivers supporting patients with Alzheimer's Disease. *Technology and Health Care, 27*(5), 531-544.

Cesario, A., D'Oria, M., Bove, F., Privitera, G., Boškoski, I., Pedicino, D., Boldrini, L., Erra, C., Loreti, C., Liuzzo, G., Crea, F., Armuzzi, A., Gasbarrini, A., Calabresi, P., Padua, L., Costamagna, G., Antonelli, M., Valentini, C., Auffray, C., & Scambia, G.

(2021). Personalized clinical phenotyping through systems medicine and artificial intelligence. *Journal of Personalized Medicine, 11*(4), 265.

Cesario, A., Simone, I., Paris, I., Boldrini, L., Orlandi, A., Franceschini, G., Lococo, F., Bria, E., Magno, S., Mulè, A., Santoro, A., Damiani, A., Bianchi, D., Picchi, D., Rasi, G., Daniele, G., Fabi, A., Sergi, P., Tortora, G., Masetti, R., Valentini, V., D'Oria, M., & Scambia, G. (2021). Development of a digital research assistant for the management of patients' enrollment in oncology clinical trials within a research hospital. *Journal of Personalized Medicine, 11*(4), 244.

Cho, C. H., & Lee, H. J. (2019). Could digital therapeutics be a game changer in psychiatry?. *Psychiatry investigation, 16*(2), 97.

Cho, G., Yim, J., Choi, Y., Ko, J., & Lee, S. H. (2019). Review of machine learning algorithms for diagnosing mental illness. *Psychiatry investigation, 16*(4), 262.

Cruz, J. A., & Wishart, D. S. (2006). Applications of machine learning in cancer prediction and prognosis. *Cancer informatics, 13*(2), 8-17.

Deparis, S., Pascale, A., Tommasi, P., & Kotoulas, S. (2018). An analytical method for multimorbidity management using Bayesian networks. *Studies in Health Technology and Informatics, 247*, 820-824.

Food and Drug Administration. (2020). Artificial intelligence and machine learning in software as a medical device. *Content current as of January, 28*, 2020. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device.

Fornecker, L. M., Muller, L., Bertrand, F., Paul, N., Pichot, A., Herbrecht, R., Chenard, M. P., Mauvieux, L., Vallat, L., Bahram, S., Cianférani, S., Carapito R., & Carapito, C. (2019). Multi-omics dataset to decipher the complexity of drug resistance in diffuse large B-cell lymphoma. *Scientific reports, 9*(1), 1-9.

Franceschini, G., Mason, E. J., Orlandi, A., D'Archi, S., Sanchez, A. M., & Masetti, R. (2021). How will artificial intelligence impact breast cancer research efficiency?. *Expert Review of Anticancer Therapy, 21*(10), 1067-1070.

Ghaheri, A., Shoar, S., Naderan, M., & Hoseini, S. S. (2015). The applications of genetic algorithms in medicine. *Oman medical journal, 30*(6), 406.

Gussoni, D. (2021). Digital therapeutics: An opportunity for Italy, and beyond – Executive Summary. *Tendenze Nuove, 4*, 3-8.

Hedman, Å. K., Hage, C., Sharma, A., Brosnan, M. J., Buckbinder, L., Gan, L. M., Shah, S. J., Linde, C. M., Donal, E., Daubert, J. C., Mälarstig, A., Ziemek, D., & Lund, L. (2020). Identification of novel pheno-groups in heart failure with preserved ejection fraction using machine learning. *Heart, 106*(5), 342-349.

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. (2018). Artificial intelligence in radiology. *Nature Reviews Cancer, 18*(8), 500-510.

Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y., & Xu, W. (2018). Applications of support vector machine (SVM) learning in cancer genomics. *Cancer genomics & proteomics, 15*(1), 41-51.

Huang, Q., Cohen, D., Komarzynski, S., Li, X. M., Innominato, P., Lévi, F., & Finkenstädt, B. (2018). Hidden Markov models for monitoring circadian rhythmicity in telemetric activity data. *Journal of The Royal Society Interface, 15*(139), 20170885.

Jain, K. K. (2009). *Textbook of Personalized Medicine*. Springer.

Jen, C. H., Wang, C. C., Jiang, B. C., Chu, Y. H., & Chen, M. S. (2012). Application of classification techniques on development an early-warning system for chronic illnesses. *Expert Systems with Applications, 39*(10), 8852-8858.

Koelzer, V. H., Sirinukunwattana, K., Rittscher, J., & Mertz, K. D. (2019). Precision immunoprofiling by image analysis and artificial intelligence. *Virchows Archiv, 474*(4), 511-522.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and structural biotechnology journal, 13*, 8-17.

Langarizadeh, M., & Moghbeli, F. (2016). Applying naive bayesian networks to disease prediction: A systematic review. *Acta Informatica Medica, 24*(5), 364.

Langlotz, C. P., Allen, B., Erickson, B. J., Kalpathy-Cramer, J., Bigelow, K., Cook, T. S., Flanders, A. E., Lungren, M. P., Mendelson, D. S., Rudie, J. D., Wang, G., Kandarpa, K., & Kandarpa, K. (2019). A roadmap for foundational research on artificial intelligence in medical imaging: From the 2018 NIH/RSNA/ACR/The Academy Workshop. *Radiology, 291*(3), 781-791.

Lee, J. (2017). Patient-specific predictive modeling using random forests: An observational study for the critically ill. *JMIR medical informatics, 5*(1), e6690.

Lenkowicz, J., Gatta, R., Masciocchi, C., Casà, C., Cellini, F., Damiani, A., Dinapoli, N., & Valentini, V. (2018). Assessing the conformity to clinical guidelines in oncology: An example for the multidisciplinary management of locally advanced colorectal cancer treatment. *Management Decision, 56*(10), 2172-2186.

Macchia, G., Ferrandina, M. G., Patarnello, S., Autorino, R., Masciocchi, C., Pisapia, V., Calvani, C., Iacomini, C., Cesario, A., Boldrini, L., Gui, B., Ruffini, V., Gambacorta, M. A., Scambia, G., & Valentini, V. (Under Review). Multidisciplinary tumor board smart virtual assistant in locally advanced cervical cancer (LACC): A proof of concept. *Journal of Translational Medicine*.

Madigan, E. A., Curet, O. L., & Zrinyi, M. (2008). Workforce analysis using data mining and linear regression to understand HIV/AIDS prevalence patterns. *Human resources for health, 6*(1), 1-6.

Marazzi, F., Tagliaferri, L., Masiello, V., Moschella, F., Colloca, G. F., Corvari, B., Sanchez, A. M., Capocchiano, N. C., Pastorino, R., Iacomini, C., Lenkowicz, J., Masciocchi, C., Patarnello, S., Franceschini, G., Gambacorta, M. A., Masetti, R., & Valentini, V. (2021). GENERATOR Breast DataMart - The novel breast cancer data discovery system for research and monitoring: Preliminary results and future perspectives. *Journal of Personalized Medicine, 11*(2), 65.

Moon, M., & Lee, S. K. (2017). Applying of decision tree analysis to risk factors associated with pressure ulcers in long-term care facilities. *Healthcare informatics research, 23*(1), 43-52.

Natanson, E. (2017). Digital therapeutics: The future of health care will be app-based. *Forbes, July 24*. https://www.forbes.com/sites/eladnatanson/2017/07/24/digital-therapeutics-the-future-of-health-care-will-be-app-based/.

National Human Genome Research Institute (2003). *What is the Human Genome Project?*. https://www.genome.gov/human-genome-project/What.

Naylor, C. D. (2018). On the prospects for a (deep) learning health care system. *Journal of the American Medical Association, 320*(11), 1099-1100.

Ngiam, K. Y., & Khor, W. (2019). Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology, 20*(5), e262-e273.

O'Bryant, S. E., Xiao, G., Barber, R., Reisch, J., Doody, R., Fairchild, T., Adams, P., Waring, S., Diaz-Arrastia, R., & Texas Alzheimer's Research Consortium. (2010). A serum protein-based algorithm for the detection of Alzheimer disease. *Archives of neurology, 67*(9), 1077-1081.

Olsen, T. G., Jackson, B. H., Feeser, T. A., Kent, M. N., Moad, J. C., Krishnamurthy, S., Lunsford, D. D., & Soans, R. E. (2018). Diagnostic performance of deep learning algorithms applied to three common diagnoses in dermatopathology. *Journal of pathology informatics, 9*(32), 1-7.

Ozsahin, I., Sekeroglu, B., Musa, M. S., Mustapha, M. T., & Uzun Ozsahin, D. (2020). Review on diagnosis of COVID-19 from chest CT images using artificial intelligence. *Computational and Mathematical Methods in Medicine*, 2020, 1-10.

Patel, S. K., George, B., & Rai, V. (2020). Artificial intelligence to decode cancer mechanism: Beyond patient stratification for precision oncology. *Frontiers in Pharmacology, 11*, 1177.

Paterson, M., & Witzleb, N. (2018). The privacy-related challenges facing medical research in an era of big data analytics: A critical analysis of australian legal and regulatory frameworks. *Journal of law and Medicine, 26*(1), 188-203.

Peiffer-Smadja, N., Rawson, T. M., Ahmad, R., Buchard, A., Georgiou, P., Lescure, F. X., Birgand, G., & Holmes, A. H. (2020). Machine learning for clinical decision support in infectious diseases: A narrative review of current applications. *Clinical Microbiology and Infection, 26*(5), 584-595.

Placidi, L., Boldrini, L., Lenkowicz, J., Manfrida, S., Gatta, R., Damiani, A., Chiesa, S., Ciellini, F., & Valentini, V. (2021). Process mining to optimize palliative patient flow in a high-volume radiotherapy department. *Technical Innovations & Patient Support in Radiation Oncology, 17*, 32-39.

Radakovich, N., Nagy, M., & Nazha, A. (2020). Artificial intelligence in hematology: Current challenges and opportunities. *Current hematologic malignancy reports, 15*(3), 203-210.

Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, K., Zhang, Y., Flores, G., Duggan, G. E., Irvine, J., Le, Q., Litsch, K., Mossin, A., Tansuwan, J., … & Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *NPJ Digital Medicine, 1*(1), 1-10.

Sangaiah, A. K., Arumugam, M., & Bian, G. B. (2020). An intelligent learning approach for improving ECG signal classification and arrhythmia analysis. *Artificial intelligence in medicine, 103*, 101788.

Schneider, A., Hommel, G., & Blettner, M. (2010). Linear regression analysis: Part 14 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International, 107*(44), 776.

Schruben, L. W. (1980). Establishing the credibility of simulations. *Simulation, 34*(3), 101-105.

Srinivas, K., Rani, B. K., & Govrdhan, A. (2010). Applications of data mining techniques in healthcare and prediction of heart attacks. *International Journal on Computer Science and Engineering, 2*(02), 250-255.

Steiner, D. F., MacDonald, R., Liu, Y., Truszkowski, P., Hipp, J. D., Gammage, C., Thng, F., Peng, L., & Stumpe, M. C. (2018). Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer. *The American journal of surgical pathology, 42*(12), 1636.

Sun, R., Limkin, E. J., Vakalopoulou, M., Dercle, L., Champiat, S., Han, S. R., Verlingue, L., Brandao, D., Lancia, A., Ammari, S., Hollebecque, A., Scoazec, J. Y., Marabelle, A., Massard, C., Soria, J. C., Robert, C., Paragios, N., Deutsch, E., & Ferté, C. (2018). A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: An imaging biomarker, retrospective multicohort study. *The Lancet Oncology, 19*(9), 1180-1191.

Sun, X., Shang, J., Wu, A., Xia, J., & Xu, F. (2020). Identification of dynamic signatures associated with smoking related squamous cell lung cancer and chronic obstructive pulmonary disease. *Journal of cellular and molecular medicine, 24*(2), 1614-1625.

Tayefi, M., Esmaeili, H., Karimian, M. S., Zadeh, A. A., Ebrahimi, M., Safarian, M., Nematye, M., Parizadeha, S. M. R., Fernsf, G. A., & Ghayour-Mobarhan, M. (2017). The application of a decision tree to establish the parameters associated with hypertension. *Computer methods and programs in biomedicine, 139*, 83-91.

Toubiana, D., Puzis, R., Wen, L., Sikron, N., Kurmanbayeva, A., Soltabayeva, A., Wilhelmi, M. M. R., Sade, N., Fait, A., Sagi, M., Blumwald, E., & Elovici, Y. (2019). Combined network analysis and machine learning allows the prediction of metabolic pathways from tomato metabolomics data. *Communications biology, 2*(1), 1-13.

Valentini, V., Cesario, A., (in press). Oltre la persona: cos'è lo "Human Digital Twin" nella medicina personalizzata. In F. Anelli, A. Cesario, M. D'Oria, C. Giuliodori, & G. Scambia (Eds.). *Persona e Medicina. Sinergie sistemiche per la medicina personalizzata.* Franco Angeli.

van Dijk, W. B., Fiolet, A. T., Schuit, E., Sammani, A., Groenhof, T. K. J., van der Graaf, R., ... & Mosterd, A. (2021). Text-mining in electronic healthcare records can be used as efficient tool for screening and data collection in cardiovascular trials: A multicenter validation study. *Journal of Clinical Epidemiology, 132*, 97-105.

Wall, M. M., & Li, R. (2009). Multiple indicator hidden Markov model with an application to medical utilization data. *Statistics in medicine, 28*(2), 293-310.

Wei, W., Visweswaran, S., & Cooper, G. F. (2011). The application of naive Bayes model averaging to predict Alzheimer's disease from genome-wide data. *Journal of the American Medical Informatics Association, 18*(4), 370-375.

Xu, W., Zhao, Y., Nian, S., Feng, L., Bai, X., Luo, X., & Luo, F. (2018). Differential analysis of disease risk assessment using binary logistic regression with different analysis strategies. *Journal of International Medical Research, 46*(9), 3656-3664.

Yi, B., Wang, G., Li, Z., Zhu, L., Li, P., Li, W., Zhi, S., Zhu, S., & Li, J., (2021). The future of robotic surgery in safe hands. *Nature Portfolio.* https://www.nature.com/articles/d42473-020-00176-y.

Zamborlini, V., Da Silveira, M., Pruski, C., Ten Teije, A., Geleijn, E., van der Leeden, M., Stuiver, M., & van Harmelen, F. (2017). Analyzing interactions on combining multiple clinical guidelines. *Artificial Intelligence in Medicine, 81*, 78-93.

Zhang, C., Peng, L., Zhang, Y., Liu, Z., Li, W., Chen, S., & Li, G. (2017). The identification of key genes and pathways in hepatocellular carcinoma by bioinformatics analysis of high-throughput data. *Medical oncology, 34*(6), 101.

Zhang, Z. (2016). Introduction to machine learning: k-nearest neighbors. *Annals of translational medicine, 4*(11).

Zhao, D., & Weng, C. (2011). Combining PubMed knowledge and EHR data to develop a weighted bayesian network for pancreatic cancer prediction. *Journal of biomedical informatics, 44*(5), 859-868.

Zhu, Z., Shen, Q., Zhu, L., & Wei, X. (2017). Identification of pivotal genes and pathways for spinal cord injury via bioinformatics analysis. *Molecular medicine reports, 16*(4), 3929-3937.

Zouri, M., Zouri, N., & Ferworn, A. (2019). An ontology approach for knowledge representation of ECG data. In *Improving Usability, Safety and Patient Outcomes with Health Information Technology* (pp. 520-525). IOS Press.

2. Of Men, Machines and Their Interactive Behavior

When AI and Robotics Meet Behavioral Economics

M.A. Maggioni, D. Rossignoli

ABSTRACT

This chapter surveys the main contributions in the behavioral economics literature dealing with the study of interactions between human and 'artificial' agents. While human-computer interactions have been extensively studied in the last decade, Artificial Intelligence and humanoid robots have entered the focus of attention of behavioral economics only very recently. Early studies used artificial agents as a 'trick' to observe a subject's reaction 'free' from other confounding factors (such as gender, ethnicity, age etc.) intrinsically embedded in confederate human partners. More recent works study the occurrence of similar or dissimilar behaviors based on the partners' nature. The experimental evidence surveyed in this chapter shows that interactions with artificial agents trigger less of an emotional response in the subjects, while subjects tend to extend to humanoid robots similar attributes that they attach to human beings, provided that robots are able to display emotions, empathy and an effective and context-related communication. The description of the main results of an original experiment, devised by the authors, concludes the paper.

Introduction

Decisions to be taken within an uncertain framework are part of our daily life as individuals and as organizations. Much of the academic research in the economic and organizational literature since World War II has been devoted to understanding the decision-making process (Bernoulli, 1954; Samuelson, 1977; Von Neumann & Morgenstern, 1944). Given the complexity of the world we live in and the limitations of the human mind – see Simon (1982) on 'limited rationality' – the task of processing a huge amount of data in order to take decisions may lead to what has been labeled 'decision stress' (Toeffler, 1970).

A previous version of this chapter had to be retracted because of insufficient references to Chugunova, M. & Sele, D. (2020). *We and It: An interdisciplinary review of the experimental evidence on human-machine interaction.* SSRN discussion paper no. 3692293. Published as Chugunova, M. & Sele, D. (2022). We and It: An Interdisciplinary Review of the Experimental Evidence on How Humans Interact with Machines. *Journal of Behavioral and Experimental Economics, 99.*

The invention of the steam engine in the XVIII century drove on the industrial revolution: harnessing the thermal energy obtained by burning coal (or wood) and heating water boosted the mechanized production of goods. Consequently, a number of phenomena took place – from the division of labor to the standardization of production, from the invention of the assembly line and the principle of mass production to the process of statistical quality control – and led to vast improvements and transformations in the production process (Brinkley, 2003; Noguchi, 1995).

About a century later, at Cambridge, Charles Babbage, founder of the Royal Astronomical (1820) and Statistical (1834) Societies, invented the *difference engine*, which many regard as the forerunner of the first 'computers'. Rapid strides in computing and data storage led to the inevitable possibility of cognition in machines: the evolution from ENIAC (Electronic Numerical Integrator and Computer) to the personal computer and from Deep Blue to AlphaGo led the way towards the creation of 'intelligent' machines.

Thus, two parallel processes accompanied the largest rise in human welfare in history (as measured by per-capita GDP): on the one hand, the use of machines to substitute human (and animal) workforces and energy for increasingly complex manual tasks (an ongoing process continually gaining momentum) and, on the other hand, the use of machines (and written instructions, codes, or software) to substitute a number of functions of the human mind.

These two processes are of great interest for economics both from a macro perspective (i.e., focused on their effects on employment, income and productivity) and a micro perspective (i.e., focused on their effects on the behavior of individual agents and/or specific industries or functions).

Indeed, for decades human interactions with computers and robots have played an important part in the production of cultural products (stories, novels, movies and, more recently, video games), populating the imagination of people across countries and cultures. In a very interesting book, Cave et al. (2020) document decades of novels, stories and fiction about artificial intelligence and how the narrative about it, a never-resolved tangle of desire and fear, has evolved over time. As the authors puts it, "However much we may hope for intelligent machines to take over the chores we hate, we never entirely put to rest the fear that they will take over everything else, too" (Cave et al., 2020, p. 189). Nowadays, these artificial agents may take very different forms (from smart-speakers to chat-bots and humanoid social robots), but ultimately their multi-faceted shapes and forms can be grouped into two large categories: 'virtual' computers, i.e. 'algorithms' that may allow for a certain

amount of autonomous 'intelligence', that we label for simplicity 'artificial intelligence' (AI), on the one hand; and 'embodied' robots with anthropomorphic features, that we label for simplicity 'social robots' (SR), on the other.

The increasing spread of interactions with computers, algorithms and robots in real life contexts, such as health care, in the marketplace or even in one's own home (Gaggioli et al., 2021; Lee et al., 2013; Schaefer et al., 2016; Schniter et al., 2020) spurred the investigation into how human people interact with robots in an attempt to answer the common underlying question, 'Can we trust artificial agents?' since trust is pivotal to sustaining economic and social interactions; as Kenneth Arrow put it, "virtually every commercial transaction has within itself an element of trust" (Arrow, 1972, p. 357). In this chapter, we will focus on a specific sub-branch of the micro perspective, the so-called behavioral economics approach, in order to present a survey of the most recent contributions in this field.

Not surprisingly behavioral economics devoted a substantial amount of effort to dealing with the determinants of trust in inter-human interactions, as documented, for instance, in the meta-analysis by Johnson and Mislin (2011), which explored the determinants of trust in 162 replications of the famous Investment Game (aka the 'Trust Game'), first developed by Berg et al. (1995), involving more than 23,000 participants. However, far less is known about how people extend trust to artificial agents, even though this field is also expanding in the behavioral economics literature. The question of trusting AI and SR is in fact becoming more and more relevant in real life, and the factors behind whether people bestow trust upon an automated agent are not obvious (Schniter et al., 2020; Taddeo et al., 2019). For this reason, behavioral economists have increasingly investigated how people behave when faced with artificial agents in experimental settings.

The spread in the use of artificial agents in economics experiments is driven by a variety of factors. Initially, in most of this literature (see, among others: Collins et al., 2016; Zanatto et al., 2019; Oliveira et al., 2020), we see the use of computers and/or robots aimed at obtaining confederate agents in multi-arm experimental settings that could be programmed according to the needs of the researcher: in this way, participants can be assigned to a number of different types of partners, ensuring that the results are not biased by any uncontrolled individual variation in the confederate agents' behavior.

More recently, we see experiments involving artificial agents extended to a more interesting question: understanding how agents behave towards an artificial partner, either a computer or an anthropomorphic robot. In this chapter, we survey the recent developments in behavioral

economics, documenting the increasing popularity of using machines in experiments, with computers, artificial intelligence and even robots serving as partners in experimental interactions.

Interacting with Computers and AI

The use of computers (intended in the broad sense as a combination of hardware and software) has largely influenced the field of economics since their inception, especially in the sub-fields of applied economics and econometrics: here the 'number crunching' power of firstly mainframe computers, then minicomputers, microcomputers and finally personal computers allowed researchers to empirically test theoretical models with increasingly large bulks of real-life data. More recently, the introduction of machine learning (ML) and other similar AI devices further allowed the use of 'algorithms' that – rather than estimating parameters within a single model chosen by the individual researcher to represent a given phenomenon – estimate and compare many alternative models.

In a sense, this approach may appear to contrast with 'economics' as a scientific discipline, where (in principle, though rarely in reality) the researcher picks a model based on principles and then estimates it once. Instead, ML and AI algorithms more radically perform model-selection tasks through a procedure which is largely, if not entirely, unsupervised and data driven.

While there are several advantages to this approach, including improved performance as well as enabling researchers to be systematic and to fully explore any possible statistical regularity hidden in the data, there are also some relevant limits in 'outsourcing' model selection to ML and AI algorithms, thus expropriating the researcher of his/her key functions.

Moreover, one should also consider that these procedures work very well when the problem is 'simple' (as in the case of model prediction and classification tasks based on criteria such as goodness of fit), but they are still suffering from severe limitations when problems are more complex and concern the search for causal inference: this is a typical situation in which there is no univocal and unbiased criterion on the basis of which to prefer one interpretation to another. In this section, however we will mainly deal with the use of computers and artificial intelligence in the sub-fields of applied game theory and behavioral economics, where they have been used to 'create' and 'move' some 'artificial agents' employed as partners in empirical experiments. In this specific sense, hardware is not really an issue.

In this chapter, therefore, we will mainly refer to 'computers' when the behavior of an 'artificial agent' is simulated by a series of pre-determined instructions (or lines of code in a software); while we will refer to AI in all the situations in which some ML algorithm is at play, thus allowing the 'artificial agent' to modify its behavior not only based on some contingency table but on the 'experience' gained in previous rounds of the game (or games).

If we refer to a recent paper on this topic by a leading scholar in the field (Camerer, 2019), the impact of AI and ML on behavioral economics is threefold: firstly AI can be used to search for new behavioral-type variables that affect choice; secondly ML techniques can be used to model certain limits on human prediction as being similar to the kinds of errors made by poorly implemented machine learning; thirdly AI and ML technologies might be able to show how firms and other institutions can both overcome and exploit human computational limits to their advantage.

In this section, we will make rather scant reference to such an authoritative suggestion, favoring instead the approach of March (2019) and other scholars, who focused on the implementation of experimental frameworks in which either AI is used to better simulate the behavior of a human partner in a given game (the subjects being either aware or unaware of the real nature of the partner itself) or, even better, the approach of Klockmann et al. (2021), whereby subjects are made aware of the AI nature of their partner in order to study the possible differential effects on human behavior when confronting a 'static' algorithm as opposed to a ML/AI one.

A starting point for considering the role of human-computer interaction (HCI) in behavioral economics lies in the fact that in 'standard' human-human strategic interactions many deviations from the predictions of theoretical models' equilibrium stem from the agents' inability to assess others' behavior. A common finding in the relative literature is that humans tend to underestimate the rationality of others or to learn too little from others' actions.

As artificial intelligence in the form of ML and AI algorithms becomes pervasive in various fields, strategic interaction between humans and artificial agents will become more and more common. Thus, it would be very interesting to discover how the results on the limits of strategic sophistication in human-human interactions translate into human-computer interactions. In other words, will humans ascribe too much or too little rationality to artificially intelligent agents?

According to a recent review of the issue, Chugunova and Sele (2020, p. 5), "a particularly striking difference between human-human and human-computer interactions is the robust finding that interacting with

automated agents triggers less of an emotional and social response in the human interaction partners. The reduced emotional response to automated agents is manifested both in terms of a less emotional immediate reaction to the agent's actions and in a generally decreased level of emotional arousal in human-computer interactions. Importantly, interacting with an automated counterpart seems to narrow the entire emotional spectrum: people react less positively to desirable actions by automated agents as well as less negatively to undesirable ones".

Thus, the increased rationality displayed in human-machine interactions is often obtained at the expenses of the socio-emotional dimension of human decisions and actions; this is well demonstrated, among others, by De Melo et al. (2016) who show that human participants tend to share less and to exploit more automated agents as compared to human ones (see Chugunova and Sele, 2020).

Will you cooperate with a computer algorithm?

A vast amount of literature in Psychology, HRI, Economics and Management surveyed in Chugunova and Sele (2020) and Nass and Moon (2000) documented that people generally (i) apply social rules, gender stereotypes and expectations to computers; (ii) react to automated feedback and reciprocated helpful acts by computers. Interestingly, as Chugunova and Sele (2020) highlight, this occurs despite people deny that computers have a personality, when directly asked.

The review of empirical studies by Chugunova and Sele (2020) show that social reactions can be induced in human subjects by both images (digital representations of humans) and verbal and/or written texts (such as chatbots or interactive computer programs).

The increasing involvement of automated agents in decision-making has attracted considerable research interest – as documented by Chugunova and Sele (2020). Several experiments, in the fields of psychology, sociology and organizational studies, were devised to investigate the following research question: 'Are humans willing to accept the involvement of automated agents in decision-making? And, if so, to what extent?'.

Interestingly, what emerges from the huge amount of research work surveyed is that, despite the fact that the research questions posed by these studies were often very similar, their findings were often utterly different.

From the review by Chugunova and Sele (2020) it emerges that some studies – such as Kantowitz et al. (1997), Dzindolet et al. (2002), Yeomans et al. (2019) – showed that humans are not willing to transfer de-

cision-making tasks to automata (algorithm aversion); others – such as Promberger and Baron (2006), Longoni et al. (2019), Logg et al. (2019) – show a preference for automated over human advice (algorithm appreciation); others – as Parasuraman and Manzey (2010), Cormier et al. (2013), Bai et al. (2021) – show a pattern of over-reliance on machines and the inability of implementing effective monitoring procedures (automation bias a/o automation induced complacency).

Trying to summarize such complex and heterogeneous literature is a rather difficult, if not impossible task. However, Chugunova and Sele (2022) suggest that:

1) interacting with automata reduces the intensity of emotional and social response in human subjects. This phenomenon has been supported both by subjective measurements (such as psychological scales) and objective behaviors (such as games' outcomes);

2) humans can sometimes have very strong reactions when a previously thought human-human interaction is revealed as an interaction with automated agents;

3) context and expectations are crucial in determining whether human subjects will be in favor or against the transfer of decision power to machines and algorithm. On the one hand humans, when the technology is easily available and presented as a complement rather than a substitute to human judgment, tend to use it intensively and to become dependent on algorithms. On the other hand, if they feel the technology as substitute or, even worse, as hierarchically superimposed to human decisions, they tend to show refusal and denial behaviors.

Playing games against computers and AI

Focusing more specifically on the game theory and behavioral economics literature, here are a number of further results.

Abric and Kahan (1972) used a Prisoner's Dilemma[1] (PD) in which subjects were told that they were playing with a human being ('another student like yourself') or an algorithm ('a programmed strategy devised by a machine') and found that participants were more cooperative toward the human being (55%) than towards the program (35%).

[1] A PD consists in a simultaneous choice by two matched agents. Both are endowed with the same initial monetary amount and have to decide whether to 'send' this amount to the partner (i.e., to cooperate) or to keep it for themselves (i.e., they do not cooperate). If both agents decide not to cooperate, their individual final payoff is equal to their initial endowment; if they both decide to cooperate, both receive a larger amount (usually twice or three-times their initial endowment); if one decides to cooperate while the other one does not, the cooperative agent receives nothing, while the other one receives an amount which is larger than the cooperative payoff.

Andreoni and Miller (1993) devised a mixed framework in which participants were asked to play a PD against a fellow human partner teamed with a computer which played according to a tit-for tat-strategy, which would, in any round, randomly (50% probability) take the role of the opponent. Despite the uncertainty involved in the framework, the addition of the computer increased the rate of cooperation displayed by subjects.

Kiesler et al. (1996) showed that allowing communication before a PD game was played increased the cooperation rate irrespective of the nature of the partner (this being either a confederate human agent or a computer program).

De Melo et al. (2011), within the framework of a negotiation task, showed that even when subjects are aware that they are playing against a computer algorithm/avatar which is able to show emotions on a face depicted on a computer screen (such as anger or happiness rather than neutral), they tended to behave as they would have if interacting with human being showing similar emotions. Similar results were obtained by Swiderska et al. (2019).

March (2019) reviewed some 90 experimental studies in which agents were partnered with a computer. His results show that humans generally act more selfishly and more rationally in the presence of computer, and they are also often able and willing to exploit these artificial players.

Finally, Klockmann et al. (2021) asked subjects to play a series of Dictator Games[2] against an AI algorithm and found that making individuals aware of the consequences of their training on the well-being of future generations changes subjects' behavior, but only when these individuals were facing the risk of being harmed themselves by future algorithmic choices.

Summarizing this last stream of empirical literature, March (2019) states that four stylized facts emerge: First, behavior often changes when human opponents are replaced by computer players. Second, subjects generically behave more selfishly and more rationally when interacting with computers. Third, people often learn to exploit computers, even

[2] In the Dictator Game, originally developed by Kahneman et al. (1986) and then designed in its current version by Forsythe et al. (1994), a person who is assigned the role of Proponent is provided with an exogenous endowment. They are matched with an anonymous partner assigned the role of Respondent, who has received no endowment. The Proponent's choice concerns how to split the endowment between themself and their partner. The Respondent has no influence over the outcome of the game. Within the standard theoretical assumptions of self-regarding agents, the Dictator Game has a unique Nash equilibrium whereby the Proponent maximizes their payoff by keeping the whole endowment, thus sending no money to the Respondent. Therefore, any deviation from the equilibrium solution in the Dictator Game is used to measure altruism and/or pure generosity.

when they possess little prior knowledge of them and when the computers do not follow a fixed strategy but are responsive to their choices (both in terms of a contingency program table or some sort of AI or ML algorithm). Fourth, sophisticated algorithms are able to outperform human subjects in certain environments.

Robots in Behavioral Experiments

The nature of inter-subject interactions has lately been extended to different types of agents by the more widespread use of robots and, specifically, of humanoid social robots. Originally, the focus and main objective of engineers was to replace a narrow range of simple manual tasks (often in manufacturing production lines) with 'artificial workers'. Only later did the range of activities expand to include more complex tasks and activities, from information to assistance, education, therapeutic mediation, and even entertainment and companionship. However, it soon became clear that, to operate in many of these fields, robots had to be able to engage in a wide range of social interactions. Furthermore, they needed to display a credible 'social presence', namely the ability to inspire in the user the 'feeling of being in the company of someone' (Gaggioli et al., 2021; Heerink et al., 2008).

Therefore, the development of social robots moved precisely in this direction. Most social robots are now, at least partially, able to display and/or perceive emotions and they can communicate using high-level dialogue, exploiting natural language patterns developed or learnt through machine learning algorithms, and making use of natural cues (such as gestures, gaze, etc.). Furthermore, as in real inter-personal interactions, social robots may be able to recognize and even learn patterns or models used by other agents; they are developed to establish and possibly maintain social relationships. The most interesting developments in the field have led to robots that are programmed to display their own unique personality and character. In short, they can learn and even develop social skills. Reliance, and hence trust, in complex technological apparatus such as robots has been an issue for a long time. Lee and See (2004) observed that automation, in general, is often problematic because of the lack of appropriate reliance. Since people tend to respond to technology socially, and usually are not able to fully grasp the complexity of automation processes, trust guides reliance on the machine and its underlying processes. Therefore, people need to trust robots in order to accept interacting with them.

Can we trust robots?

A number of studies, especially in psychology, investigated whether people feel some level of trust in robots or not. There is a strong relationship between trust in human-robot interaction (HRI) and trust in automation in general and the latter has already been studied especially by engineers and related scholars (e.g., Chen et al., 2010; Lee & See, 2004; Parasuraman et al., 2008). However, as Hancock et al. (2011, p. 518) observe, robots differ from other automated systems because they "are sometimes built in a fashion that approximates human or animal form". Therefore, it could be argued that trust towards robots in HRI contexts follows different patterns from trust in HCI or interactions between humans and automation in general. Hancock et al. (2011) and Schaefer et al. (2016) provide two meta-analyses showing mixed evidence about how trust in robots is affected by specific design features or by human related aspects. In general, it seems that people tend to trust robots more when they present more 'human-like' characteristics, but, as noted by Schniter et al. (2020), the results provided by these two meta-analyses (especially the first one by Hancock et al., 2011) may be problematic since only a limited number of studies employ objective measures of trust (such as the Trust Game proposed by Berg et al., 1995), almost exclusively relying on subjective self-administered questionnaires[3]. This kind of questionnaire can contribute to unreliable responses due to deception by participants, who may declare that they trust a robot while their actual behavior, especially in a real life situation, shows otherwise.

Recent empirical research in HRI has implemented experimental frameworks in which human subjects have been partnered with humanoid robots in social dilemmas. These studies, relying on objective measures of trust or cooperation, by making use of incentivized tasks, investigated more neatly the potential determinants of human behavior towards robots, by manipulating some features of the robot partner or of the experimental setting, or both.

The most recent account of trust in HRI is provided by Schniter et al. (2020). In their study, the authors explore trust in HRI by comparing the outcome of a Trust Game (Berg et al., 1995) when people are matched with a fellow human or a robot (in this case a mere algorithm) which mimics the behavior in a Trust Game of human subjects. Their starting point is the assumption that since humans and robots are very

[3] It is also worth noting that these two meta-analyses mainly focused on military applications of robots i.e., high-risk situations in which relying on a robot, and complying, for instance, with its instructions or suggestions, may be crucial to saving the lives of troops or civilians in war-like scenarios.

different in many aspects, differences should arise in people's behavior towards real human partners and robots. Indeed, their results show no difference across human and robot conditions, suggesting that in all experimental conditions people were focused on a simple economic task aimed at gaining money and information about their partners' behavior. However, they also add "While initial trust in humans versus robots may not differ, we have reason to expect that certain emotional responses could – and that these emotional reactions to trust in humans and robots might bring about their own effects if given the opportunity" (Schniter et al., 2020, p. 22). In fact, in Schniter et al. (2020), robots are not "fully embodied", and the authors acknowledge this as being a strong limitation of their study. A crucial aspect of modern robotics, especially social robotics, is the attempt to create automated robots that resemble human beings. In other words, anthropomorphism is a key issue in understanding the ability of people to trust robots. At present, a variety of humanoid robots have been made available for experiments, presenting different extents of autonomy, and very different sizes, shapes and levels of anthropomorphism. Among the wide range of different robots, the use of NAO has become increasingly popular due to the robot's features and capabilities, which make it appropriate for experimental (especially clinical) research. NAO is a small humanoid robot produced by Softbank Robotics (see Gelin, 2018, for references) that has become very popular in social sciences due to its flexibility and user-friendliness. It resembles a sort of 'puppet', but its autonomous routine gives a visible impression of a living thing. Robaczewski et al. (2021) document 70 experimental studies in which people are involved in a HRI with NAO, but only a few are actual behavioral economics experiment, investigating trust and cooperation in HRIs. In particular, 26 studies are specifically designed to study social interactions between human and robots, and, to the best of our knowledge, only one of these studies (i.e., Sandoval et al., 2016) ask participants to play a PD – and an Ultimatum Game (UG).

EMBODIED ROBOTS. In general, only a few studies directly investigate trust (or more generally cooperation that indirectly requires some form of trust towards the partner) towards embodied humanoid robots (both using NAO or other devices). In one of the most relevant, DeSteno et al. (2012) set up an experiment to investigate how non-verbal cues affect co-operation. They allow participants to play a Give-Some-Game (GSG)[4], being matched with a human partner in either a face-to-face or

[4] A GSG is very similar to the PD, the only difference being that while in the traditional PD the decision about cooperation involves all the initial endowment as if it were a single 'token', in a GSG agents can decide to send (or keep) part of all of their endowment.

web-based interaction, to obtain a baseline value of cooperation rates. In a second experiment, half of the agents are matched with a humanoid robot, instead of a human partner, that has been programmed in order to present non-verbal cues associated with a less trustworthy behavior. The presence of these signals makes respondents less prone to cooperate when facing robots, suggesting that the human mind responds to trust-relevant signals displayed by humanoid robots in the same manner as they respond to similar signals displayed by humans. This finding corroborates the results obtained by De Melo et al. (2011) in the context of human-computer interaction in which expressions of anger were attributed to an avatar rather than performed by a proper humanoid robot. The work by DeSteno et al. (2012) provides some of the first evidence that patterns of cooperation in HRI follow analogous rules to those of human-human interaction (HHI). More recently, Zanatto et al. (2019) found similar results in an experimental setting in which participants on average rewarded cooperation with cooperation, while punishing selfishness with selfishness. In this study, by manipulating the specific profiles of the robots, the author also show that cooperation is increased when robots present more human-like characteristics, but only when the payoff at stake is lower. Sandoval et al. (2016) develop an experiment in which participants are matched with either a human or a humanoid robot (in this case NAO) to play both a PD and an UG. The results provide initial evidence of some differences in the behavior of people when faced with a robot or a fellow human being. Participants in fact collaborated more with humans than with a robot, although they tended to be equally reciprocal with both agents. Interestingly, Sandoval et al. (2016) also collected information, through self-administered subjective questionnaires, on emotional perceptions of human and robotic partners. While robots were perceived as less open and agreeable than humans, they found no difference in terms of perceived consciousness, extroversion and emotional stability[5].

ANTHROPOMORPHISM. It is well known in the literature on robotics that as people interacting with robots are increasingly unlikely to be technically trained experts, they are more likely to use random intuitive approaches to this interaction. This process may lead to a phenomenon

[5] It is worth noting that the experiment set up by Sandoval et al. (2016) may display some limitations. In particular, human partners were asked to be neutral, to interact as little as possible with participants and to avoid conversations. They were even instructed to nod at participants in response to their greetings at the beginning of the experiment. Moreover, an experimenter (called a 'referee') was always present during the interactions in the lab room, thus likely introducing a strong bias in the participants' behavior.

known as Uncanny Valley (Mori, 1970), i.e., the occurrence of "unexpectedly negative reactions to imperfectly human robots" (Mathur & Reichling, 2016, p. 1). More specifically, the likeability of a socially interactive robot tends to increase as its human resemblance increases, but it drops when the appearance of the robot presents minor but clearly detectable imperfections, making it more 'creepy' than likeable. Therefore, modelling the appearance of a robot may play a crucial role in the likelihood of people trusting and engaging in cooperative behaviors with them. A key question in behavioral economics is whether the degree of anthropomorphism – which has been proved to activate the human subjects' social brain areas, consequently increasing the perception of robots as social partners (for a survey, see Manzi et al. 2021a) – is linked to cooperation and trust levels. Krach et al. (2008) set up an experiment in which participants had to play a version of the traditional PD against four opponents: a human partner (HP), an anthropomorphic robot (AR), a functional robot (FR), represented by a simple Lego-robot and a computer partner (CP). The authors intended these four alternatives to represent decreasing levels of human-likeness. Interestingly, in this experiment no difference occurs in the behavioral patterns (i.e., all participants cooperate about 60% of the times irrespective of the type of the partners). However, through a set of questionnaires, the authors were also able to gather information about several dimensions of the interaction, reporting that participants on average considered the FR (i.e., the less anthropomorphic robot) to be 'dumber', while expressing greater pleasure in interacting with human-like robots.

EMOTIONS, EMPATHY, AND SOCIAL INTERACTIONS. In an attempt to bridge the Uncanny Valley and provide increasingly useful and agreeable devices, robotics has made giant leaps in the last decades, with robots that we can perceive as autonomous selves, generating emotions and even feelings in the interlocutor. In a meta-analysis focusing on the most popular of these robots in the social sciences, namely NAO, Robaczewski et al. (2021) studied how this device may be helpful on several levels. Among them, the authors show first of all that it is crucial to understand which factors heighten NAO's usability in social interactions. Empirical evidence demonstrates NAO's effectiveness in influencing people's behavior. Furthermore, NAO's ability to engage in social interactions allows users to feel more comfortable with the robot, increasing both likability and their positive attitude towards it. Second, Robaczewski et al. (2021) show that people tend to be more likely to interact with a robot which is perceived as empathic: people like it when robots can understand their emotions and express emotion through gestures. Overall, the survey by

Robaczewski et al. (2021) shows that HRI, in the limited case of NAO, can be encouraged when robots are perceived as being as similar as possible to human beings, not necessarily in terms of shape and appearance (which may lead to the unfavorable Uncanny Valley situation), rather in terms of being able to express human-like characteristics. Manzi et al. (2021b) compared the expectations and attribution of mental state (AMS) among two groups of university students when faced with two different robots (NAO and Pepper, both developed by Softbank Robotics. The results showed that both the observation of interaction and the physical appearance of the robots affect AMS, with greater AMS to Pepper than to NAO. People's expectations are influenced by the interaction but are independent of the type of robot.

COMMUNICATION. Taken together, all the above suggests that the complexity of the task as well as the profile of the humanoid robots seem to be related to more cooperative behaviors in HRI, although the pattern is not clear-cut, and some heterogeneity is observed. Behavioral economics has so far shown that people may interact with robots as if they were (almost) human beings, provided they are able to display essential features common to human beings, such as emotion, empathy and, finally, communication. The above-mentioned meta-analysis by Robaczewski et al. (2021) has also shown that the way the robot is able to communicate is essential when investigating the social interactions between a human and a robot. Building on these premises, Maggioni and Rossignoli (2021) set up a new experiment to investigate how cooperative behavior differs when respondents are faced with a humanoid robotic partner as compared to a fellow human being, taking into account precisely the effect of verbal communication. The experimental design is composed of two distinct phases. In phase 1, the subjects (university students) are asked to answer an online questionnaire and to play an incentivized task (a PD) against an unknown, anonymous partner. At the end of phase 1 subjects are asked whether they want to proceed to phase 2 in the university lab. At this point, subjects would be able to find out the result of the interaction, be rewarded and possibly have further interactions with their partner. The crucial feature of the experimental design is the randomized allocation of partners. In fact, in phase 2 subjects were randomly matched with either a human (H) or robot (R) partner in the interactive situation (the PD); furthermore, half of them were randomly administered a stimulus (treatment) under both the H and R experimental conditions. The treatment consisted of a 'dialogic verbal reaction' (DVR) that the partner delivers after observing a sub-optimal aggregate outcome of the interaction. Different stimuli were administered

depending on the observed outcomes in the PD[6] to gently remind participants that cooperation is a 'superior' outcome. The preliminary results of this study, involving more than 300 subjects, show that while on average respondents tend to cooperate more with humans rather than humanoid robots, when DVR is administered this difference disappears. In other words, this result seems to suggest that when robots are made able to react verbally to an unfavorable outcome in the PD, people tend to react in a similar way as with human beings, suggesting that verbal cues may help cooperation in HRI.

Conclusion

While engineers and psychologists have been studying human-computer and human-robot interactions over the last decade – in the sub-fields of social robotics and human-robot interactions – AI and humanoid robots have entered the space of behavioral economics only very recently. Indeed, use of these devices has become common not only in investigating the economic behavior of subjects by exploiting their flexibility and usability as artificial confederate partners, but more recently in specifically investigating how human subjects perceive computer and robotic partners when engaged in (mostly) strategic interactions with them. On the one hand, the use of AI and ML has been useful for providing new assessments of rationality and investigating how much rationality people attribute to computers. On the other hand, when automated algorithms are embodied in humanoid devices, behavioral economists have engaged in investigating whether people are able to engage in social interactions with robots, especially in terms of trust and cooperative behavior.

The main takeaways of this surveys are as follows:

1) "interacting with automated agents triggers a reduced emotional and social response". (Chugunova and Sele, 2022, p. 4).

The decrease in emotional response may have both positive (increased rationality) and negative (reduced empathy and cooperation) effects. Thus the net effect is crucially determined by the context in which human decisions and interactions take place (Chugunova and Sele, 2020);

2) in human-robot interaction, people tend to extend attributes to humanoid robots that are similar to those they attach to human beings, provided the robots are able to display emotions (either through non-

[6] No DVR is activated when the aggregate Pareto optimal outcome is obtained, i.e., when both subject and partner cooperate, since the DVR aim is to move the interaction toward the social optimum.

verbal or verbal cues), possibly demonstrate a certain degree of empathy and, importantly, communicate efficiently. The degree of anthropomorphism seems less important in determining the propensity to engage in cooperative behavior or extend trust to a humanoid robot. More than its shape and physical appearance, it is the capacity to display emotions and communicate that induces people to treat robots (almost) as they were human.

As both algorithms and robots will continuously improve their performance and reduce the perceived distance between human and artificial agents, we think these streams of research should continuously improve and proceed for two reasons: on the one hand to keep up with technological advancement in the field; on the other, to foster critical thinking and understanding of the increasing role that algorithms and machines will play in an imperfect and contradictory human environment, with a series of problematic psychological, social, economic, philosophical and even moral consequences that at present can perhaps only be envisaged.

In the words of the novel *Machines Like Me* by Ian McEwan, these consequences are expressed as follows: "We create a machine with intelligence and self-awareness and push it out into our imperfect world. Devised along generally rational lines, well disposed to others, such a mind soon finds itself in a hurricane of contradictions" (McEwan, 2019, p. 149). In order to avoid the gloomy conclusion of the novel we need to think well in advance. The research documented in this paper and in the book that you are currently reading is, in our opinion, a step in the right direction.

References

Abric, J.C., & Kahan, J.P. (1972). The effects of representations and behavior in experimental games. *European Journal of Social Psychology, 2* (2), 129-144.

Andreoni, J., & Miller, J. H. (1993). Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. *The Economic Journal 103*(418), 570-585.

Arrow, K. J. (1972). Gifts and exchanges. *Philosophy & Public Affairs 1*(4), 343-362.

Bai, B., Dai, H., Zhang, D., Zhang, F., and Hu, H. (2021). The impacts of algorithmic work assignment on fairness perceptions and productivity. In *Academy of Management Proceedings: Volume 2021* (pp. 12335). Academy of Management Briarcliff Manor.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior 10*(1), 122-142.

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. *Econometrica 22*(1), 23-36.

Brinkley, D. (2003). *Wheels for the World: Henry Ford, his Company, and a Century of Progress, 1903-2003*. Viking.

Camerer, C. F. (2019). Artificial intelligence and behavioral economics. In A. Agrawal, J. Gans, & A. Goldfarb, *The Economics of Artificial Intelligence* (pp. 587-610). University of Chicago Press.

Cave, S., Dihal, K., & Dillon, S. (2020). *AI narratives: A history of Imaginative Thinking about Intelligent Machines*. Oxford University Press.

Chen, J. Y., Barnes, M. J., & Harper-Sciarini, M., (2010). Supervisory control of multiple robots: Human-performance issues and user-interface design. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 41*(4), 435-454.

Chugunova, M. & Sele, D. (2020). We and it: An interdisciplinary review of the experimental evidence on human-machine interaction. *Max Planck Institute for Innovation & Competition Research Paper* (20-15).

Chugunova, M. & Sele, D. (2022). We and it: An Interdisciplinary Review of the Experimental Evidence on How Humans Interact with Machines. *Journal of Behavioral and Experimental Economics, 99*.

Collins, M. G., Juvina, I., & Gluck, K. A. (2016). Cognitive model of trust dynamics predicts human behavior within and between two games of strategic interaction with computerized confederate agents. *Frontiers in Psychology, 7*, 49.

Cormier, D., Newman, G., Nakane, M., Young, J. E., & Durocher, S. (2013). Would you do as a robot commands? An obedience study for human-robot interaction. In *International Conference on Human-Agent Interaction*.

De Melo, C., Marsella, S., & Gratch, J. (2016). People do not feel guilty about exploiting machines. *ACM Transactions on Computer-Human Interaction (TOCHI), 23*(2), 1-17.

DeSteno, D., Breazeal, C., Frank, R. H., Pizarro, D., Baumann, J., Dickens, L., & Lee, J. J. (2012). Detecting the trustworthiness of novel partners in economic exchange. *Psychological Science 23*(12), 1549-1556.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (2002). The perceived utility of human and automated aids in a visual detection task. *Human Factors 44*(1), 79-94.

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior 6*(3), 347-369.

Gaggioli, A., Chirico, A., Di Lernia, D., Maggioni, M. A., Manzi, F., Marchetti, A., Massaro, D., Rossignoli, D., Sandini, G., Villani, D., Riva, G., & Sciutti, A. (2021). Machines like us and people like you: Toward human-robot shared experience. *Cyberpsychology, Behavior, and Social Networking 24*(5), 357-361.

Gelin, R. (2018). *NAO*. Springer.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors 53*(5), 517-527.

Heerink, M., Kröse, B., Evers, V., & Wielinga, B. (2008). The influence of social presence on acceptance of a companion robot by older people. *Journal of Physical Agents, 2*, 33-40.

Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology 32*(5), 865-889.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *The American Economic Review 76*(4), 728-741.

Kantowitz, B. H., Hanowski, R. J., & Kantowitz, S. C. (1997). Driver acceptance of unreliable traffic information in familiar and unfamiliar settings. *Human Factors 39*(2), 164-176.

Kiesler, S., Sproull, L., & Waters, K. (1996). A prisoner's dilemma experiment on cooperation with people and human-like computers. *Journal of Personality and Social Psychology 70*(1), 47.

Klockmann, V., von Schenk, A., & Villeval, M. C. (2021). *Artificial Intelligence, Ethics, and Intergenerational Responsibility*. GATE Working Paper Series.

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS One 3*(7), e2597.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors 46*(1), 50-80.

Lee, J. J., Knox, B., Baumann, J., Breazeal, C., & DeSteno, D. (2013). Computationally modeling interpersonal trust. *Frontiers in Psychology, 4*, 893.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes, 151*, 90-103.

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research 46*(4), 629-650.

Maggioni, M. A., & Rossignoli, D. (2021). If it looks like a human and speaks like a human ... dialogue and cooperation in human-robot interactions. arXiv.org.

Manzi F., Di Dio, C., Di Lernia, D., Rossignoli, D., Maggioni, M.A., Massaro, D., Marchetti A., & Riva, G. (2021a). Can you activate me? From robots to human brain. *Frontiers in Robotics and AI, 08*, 633514.

Manzi F., Massaro, D., Di Lernia, D., Maggioni, M.A., Riva, G., & Marchetti, A. (2021b). Robots are not all the same: Young adults' expectations, attitudes, and mental attribution to two humanoid social robots. *Cyberpsychology, Behavior, and Social Networking, 24*(5), 307-314.

March, C. (2019). *The Behavioral Economics of Artificial Intelligence: Lessons from Experiments with Computer Players*. Number 154. BERG Working Paper Series.

Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the uncanny valley. *Cognition, 146*, 22-32.

McEwan, I. (2019). *Machines Like Me: And People Like You*. Jonathan Cape.

Mori, M. (1970). The uncanny valley. *Energy 7*, 33-35.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues 56*(1), 81-103.

Noguchi, J. (1995). The legacy of W. Edwards Deming. *Quality Progress, 28*(12), 35.

Oliveira, R., Arriaga, P., Santos, F. P., Mascarenhas, S., & Paiva, A. (2020). Towards prosocial design: A scoping review of the use of robots and virtual agents to trigger prosocial behaviour. *Computers in Human Behavior, 114*, 106547.

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors, 52*(3), 381-410.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making, 2*(2), 140-160.

Promberger, M., & Baron, J. (2006). Do patients trust computers? *Journal of Behavioral Decision Making, 19*(5), 455-468.

Robaczewski, A., Bouchard, J., Bouchard, K., & Gaboury, S. (2021). Socially assistive robots: The specific case of the NAO. *International Journal of Social Robotics*, 13, 795-831

Samuelson, P. A. (1977). St. Petersburg paradoxes: Defanged, dissected, and historically described. *Journal of Economic Literature, 15*(1), 24-55.

Sandoval, E. B., Brandstetter, J., Obaid, M., & Bartneck, C. (2016). Reciprocity in human-robot interaction: A quantitative approach through the prisoner's dilemma and the ultimatum game. *International Journal of Social Robotics, 8*(2), 303-317.

Schaefer, K. E., Chen, J. Y., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors, 58*(3), 377-400.

Schniter, E., Shields, T. W., & Sznycer, D. (2020). Trust in humans and robots: Economically similar but emotionally different. *Journal of Economic Psychology, 78*, 102253.

Simon, H. A. (1982). *Models of Bounded Rationality*. MIT Press.

Swiderska, A., Krumhuber, E. G., & Kappas, A. (2019). Behavioral and physiological responses to computers in the ultimatum game. *International Journal of Technology and Human Interaction (IJTHI), 15*(1), 33-45.

Taddeo, M., McCutcheon, T., & Floridi, L. (2019). Trusting artificial intelligence in cybersecurity is a double-edged sword. *Nature Machine Intelligence, 1*(12), 557-560.

Toeffler, A. (1970). *Future Shock*. Random House.

Von Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press.

Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making, 32*(4), 403-414.

Zanatto, D., Patacchiola, M., Goslin, J., & Cangelosi, A. (2019). Investigating cooperation with robotic peers. *PLoS One, 14* (11), e0225028.

3. Can You Activate Me?

From Robots to Human Brain

F. Manzi, C. Di Dio, D. Di Lernia, D. Rossignoli, M.A. Maggioni, D. Massaro, A. Marchetti, G. Riva

ABSTRACT

Humans tend to anthropomorphize social robots, but experiencing their limits repositions their ontological status. Recently, behavioural studies in the field of human-robot interaction have shown that robots are perceived as plausible human partners in different contexts. Also, studies combining neuroscience, cognitive science and robotics generally inform us that our brain responds quite similarly when stimuli involve both human and robotic agents, as long as the robot's visible behaviour (i.e., movements and emotional expressions) resembles the human one. This would activate motor and emotional resonance mechanisms. However, our cognitive and control processes prevail over the tendency to anthropomorphize the robots after experiencing their limits in real interactions, thus diminishing the response of automatic systems through a top-down process.

From Robots to Human Brain

The effectiveness of social robots has been widely recognized in different contexts of humans' daily life, but still little is known about the brain areas activated by observing or interacting with a robot. Research combining neuroscience, cognitive science and robotics can provide new insights into both the functioning of our brain and the implementation of robots. Behavioural studies on social robots have shown that the social perception of robots is influenced by at least two factors: physical appearance and behavior (Marchetti et al., 2018). How can neuroscience

---

This chapter was originally published as Manzi, F., Di Dio, C., Di Lernia, D., Rossignoli, D., Maggioni, M.A., Massaro, D., Marchetti, A., & Riva, G. (2021). Can you activate me? from robots to human brain. *Frontiers in Robotics and AI, 8*, 14. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. FM conceived the idea. FM, DDL and DR selected the articles. FM, CD, and GR finalised the idea. All authors contributed to the article. This work has been supported by Università Cattolica del Sacro Cuore, Milan (D3.2 - 2018 - Human-Robot Confluence project).

explain such findings? To date, studies have been conducted through the use of both EEG and fMRI techniques to investigate the brain areas involved in human-robot interaction. These studies have mainly addressed brain activations in response to paradigms involving either action performance or charged of an emotional component (Figure 1).

A first set of studies analysed the effect of different types of robots varying in their level of physical anthropomorphism on the activation of the Mirror Neuron Mechanism (MNM). The neuronal activities examined through fMRI indicated that the activation of medial premotor cortex (MPFC) increased linearly over the degree of human-likeness of the robots, from the most mechanical to android ones (Krach et al., 2008). Electroencephalography (EEG) data associated with the mu wave – related to the MNM – showed a modulation of the mu rhythm as a function of the robotic agent resemblance to the human (Urgen et al., 2013; Matsuda et al., 2016). Furthermore, the fMRI findings on MNM indicated that the premotor cortex is similarly activated when actions are performed by different types of robots (more mechanical or android) (Saygin et al., 2012).

These evidences support the hypothesis that the premotor cortex is 'automatically' triggered in response to both simple and complex goal-directed and intentional actions, revealing a sensitivity to both the living and non-living ontological status of the agent (Gazzola et al., 2007; Saygin et al., 2012). Activation of the premotor cortex was also found in response to a human or robotic face expressing emotions (Chaminade et al., 2010). Several studies in humans have found that the premotor cortex is involved in the process of emotion recognition by encoding the motor pattern, (i.e., facial expression) that characterizes a given emotional state. The visuo-motor information processed in premotor cor-

tex is translated into affective information by means of the insula that acts as a relay station between the cortical and subcortical areas, such as the amygdala, involved in processing emotional stimuli (e.g., Carr et al., 2003; Wicker et al., 2003; Iacoboni, 2009). Likewise, the parieto-prefrontal network characterizing the MNM has been found to be particularly sensitive to biological movement (e.g., Dayan et al., 2007; Casile et al., 2009; Di Dio et al., 2013). Accordingly, it was demonstrated that observing a motor or emotional behaviour performed by a human-like robotic agent, resembling the human kinematics, may be sufficient to activate MNM (Gazzola et al., 2007; Chaminade et al., 2010). Additionally, investigating the vitality forms of movement, which characterize the style of an action, (e.g., rude *vs.* gentle) (Stern, 1985, 2010), it was shown that, besides the activation of the MNM, vitality forms activate also the dorso-central insular cortex (Di Dio et al., 2013; Di Cesare et al., 2016), which represents the relay through which information about the action style (i.e., action kinematics) processed in the parietal MNM is invested with an affective quality. Most importantly, very recent neuroscientist evidence has shown that the same brain areas whose activation is stimulated by human vitality forms can be also evoked by robots' actions performed by simulating human kinematics (Di Cesare et al., 2020), thus conveying information about the robot's 'emotion state'.

However, the activation of other brain areas besides the MNM, such as ventral visual areas, may be required to accommodate the robot's inconsistent kinematics associated with simple *vs.* complex goal-directed actions (Gazzola et al., 2007). Similarly, fMRI data showed a greater activation of posterior occipital and temporal visual cortices in response to facial expression of robot emotions compared to human emotions, reflecting a further level of processing in response to the unfamiliar stimulus (i.e., the face of the robot) (Chaminade et al., 2010; Jung et al., 2016). Additionally, the increase in frontal theta activity – associated with the recovery from long-term memory – measured through EEG is greater for a mechanical robot than a human or android (Urgen et al., 2013), highlighting once more the involvement of a compensation process for the analysis of robot stimuli. More specifically, this finding indicates that a lower level of physical robot anthropomorphism requires more resources from memory systems to bridge the semantic gap between the agent and its action (Urgen et al., 2013). People's sense of affiliation with a robot during interactions is at least partially explained by the emotional responses to the robot's behaviour. Still few studies have analysed the brain activation in response to the emotions expressed by robots. EEG data suggest that people can recognize the bodily emotions expressed by a robot, including joy and sadness, although not all the expressed emotions elicit a significant brain response in the viewer

(Guo et al., 2019). Additionally, also fMRI data indicate that emotional expressions (i.e., joy, anger and disgust) are perceived as more emotional when expressed by a human face than by a robot (Chaminade et al., 2010). As argued above, these differences could be explained by a non-perfect alignment between the robot and human kinematics expressing the emotional quality of movement.

Additional studies had investigated neural activation patterns related to emotional reactions when people observe a robot or a human into a violent situation. The fMRI data showed no differences in activation patterns in areas of emotional resonance when a violent action was experienced by a human or robot (Rosenthal-von der Pütten et al., 2014).

Suzuki et al. (2015) found a similar brain response measured through EEG when observing images showing either a finger of a robotic hand or a human hand getting cut with a scissor. In particular, the authors found an increased neural response in the ascending phase (i.e., 350-500 ms after stimulus onset) of the P3 component at the frontal-central electrodes by painful human stimuli but not painful robot stimuli, although the difference was only marginal; in contrast, no differences were found for empathy directed toward humans and robots in the descending phase of P3 (i.e., 500-650 ms after stimulus onset). Based on these results, the authors suggest that humanity of the observed agent (human *vs.* robot) partially modulates the top-down controlled processes of empathy for pain (Suzuki et al., 2015), possibly also due to a greater difficulty in taking the robot's perspective compared to the human one (Suzuki et al., 2015). In this context, it is important to underline that these pioneering studies on empathy are quite heterogeneous with respect to both the techniques adopted, and the stimuli used, which vary greatly both in terms of the type of robotic agent and experimental paradigm. To sum up, our brain systems respond in an 'embodied' fashion to the observation of experimental conditions involving the actions of a robot with biological or semi-biological dynamics. However, we suggest that this effect is only transitory or anyway limited to experimental settings. Our consideration is supported by the results by Cross et al. (2019) indicating that a period of real interaction with a social robot can disambiguate its ontological status, thus repositioning the robot in the 'non-living category'. This may be plausibly explained by the activation of top-down cognitive mechanisms that regulate the activity of our brain and that highlight the emerge of differences between the brain response to the human *vs.* robot stimuli. In other words, the automatic activation of embodied mechanisms mediated by the MNM when we observe a robot performing actions or experiencing particular human-like emotional states (e.g., violence or pain) are facilitated in a first 'encounter' with the robot, also given our natural tendency to anthro-

pomorphize many different entities. Prior experience with the robot's actual physical and psychological limits, on the other hand, provides us with a contextual frame of reference whereby top-down processes would modulate or inhibit the response of automatic mechanisms (Paetzel et al., 2020; Rossi et al., 2020). Concluding, although further studies are necessary, we can state that the level of physical anthropomorphism, the type and kinematics of the actions performed by robots jointly activate the social brain areas, consequently increasing the perception of robots as social partners. The use of additional techniques such as Virtual Reality could also prove effective in this respect (Riva et al., 2018; Riva et al., 2019).

References

Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the national Academy of Sciences, 100*(9), 5497-5502.

Casile, A., Dayan, E., Caggiano, V., Hendler, T., Flash, T., & Giese, M. A. (2010). Neuronal encoding of human kinematic invariants during action observation. *Cerebral Cortex, 20*(7), 1647-1655.

Chaminade, T., Zecca, M., Blakemore, S. J., Takanishi, A., Frith, C. D., Micera, S., Dario, P., Rizzolatti, G., Gallese, V., & Umiltà, M. A. (2010). Brain response to a humanoid robot in areas implicated in the perception of human emotional gestures. *PloS One, 5*(7), e11577.

Cross, E. S., Hortensius, R., & Wykowska, A. (2019). From social brains to social robots: Applying neurocognitive insights to human-robot interaction. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 374*(1771), 20180024.

Dayan, E., Casile, A., Levit-Binnun, N., Giese, M. A., Hendler, T., & Flash, T. (2007). Neural representations of kinematic laws of motion: Evidence for action-perception coupling. *Proceedings of the National Academy of Sciences, 104*(51), 20582-20587.

Di Cesare, G., Valente, G., Di Dio, C., Ruffaldi, E., Bergamasco, M., Goebel, R., & Rizzolatti, G. (2016). Vitality forms processing in the insula during action observation: A multivoxel pattern analysis. *Frontiers in human neuroscience, 10*, 267.

Di Cesare, G., Vannucci, F., Rea, F., Sciutti, A., & Sandini, G. (2020). How attitudes generated by humanoid robots shape human brain activity. *Scientific Reports, 10*(1), 16928.

Di Dio, C., Di Cesare, G., Higuchi, S., Roberts, N., Vogt, S., & Rizzolatti, G. (2013). The neural correlates of velocity processing during the observation of a biological effector in the parietal and premotor cortex. *Neuroimage, 64*, 425-436.

Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007). The anthropomorphic brain: The mirror neuron system responds to human and robotic actions. *Neuroimage 35*(4), 1674-1684.

Guo, F., Li, M., Qu, Q., & Duffy, V. G. (2019). The effect of a humanoid robot's emotional behaviors on users' emotional responses: Evidence from pupillometry and electroencephalography measures. *International Journal of Human-Computer Interaction, 35*(20), 1947-1959.

Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual review of psychology, 60*, 653-670.

Jung, C. E., Strother, L., Feil-Seifer, D. J., & Hutsler, J. J. (2016). Atypical asymmetry for processing human and robot faces in autism revealed by fNIRS. *PloS One 11*(7), e0158804.

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PloS One, 3*(7), e2597.

Marchetti, A., Manzi, F., Itakura, S., & Massaro, D. (2018). Theory of mind and humanoid robots from a lifespan perspective. *Zeitschrift Für Psychologie 226*(2), 98-109.

Matsuda, G., Hiraki, K., Ishiguro, H., Matsuda, G., Hiraki, K., & Ishiguro, H. (2016). EEG-based mu rhythm suppression to measure the effects of appearance and motion on perceived human likeness of a robot. *International Journal of Human-Computer Interaction*, *5*(1), 68-81.

Paetzel, M., Perugia, G., & Castellano, G. (2020). The persistence of first impressions: The effect of repeated interactions on the perception of a social robot. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 73-82). IEEE.

Riva, G., Wiederhold, B. K., Chirico, A., Di Lernia, D., Mantovani, F., & Gaggioli, A. (2018). Brain and virtual reality: What do they have in common and how to exploit their potential. *Annual Review of CyberTherapy and Telemedicine, 16*, 3-7.

Riva, G., Wiederhold, B. K., Di Lernia, D., Chirico, A., Riva, E. F. M., Mantovani, F., Cipresso, P., & Gaggioli, A. (2019). Virtual reality meets artificial intelligence: The emergence of advanced digital therapeutics and digital biomarkers. *Annual Review of CyberTherapy and Telemedicine, 17*, 3-7.

Rosenthal-von der Pütten, A. M., Schulte, F. P., Eimler, S. C., Sobieraj, S., Hoffmann, L., Maderwald, S., Brand, M., & Krämer, N. C. (2014). Investigations on empathy towards humans and robots using fMRI. *Computers in Human Behavior, 33*, 201-212.

Rossi, A., Dautenhahn, K., Koay, K. L., Walters, M. L., & Holthaus, P. (2020). Evaluating people's perceptions of trust in a robot in a repeated interactions study. In A. R. Wagner, D. Feil-Seifer, K. S. Haring, S. Rossi, T. Williams, H. He, et al. (Eds.), *International Conference on Social Robotics* (pp. 453-465). Springer.

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2012). The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social cognitive and affective neuroscience, 7*(4), 413-422.

Stern, D. N. (2010). *Forms of Vitality*. Oxford University Press.

Stern, D. N. (1985). *The Interpersonal World of the Infant*. Basic Books.

Suzuki, Y., Galli, L., Ikeda, A., Itakura, S., & Kitazaki, M. (2015). Measuring empa-

thy for human and robot hand pain using electroencephalography. *Scientific Reports, 5*(1), 15924.

Urgen, B. A., Plank, M., Ishiguro, H., Poizner, H., & Saygin, A. P. (2013). EEG theta and Mu oscillations during perception of human and robot actions. *Frontiers in Neurorobotics, 7*, 19.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in My insula: The common neural basis of seeing and feeling disgust. *Neuron, 40*(3), 655-664.

4. A Media-Studies Take on Social Robots as Media-Machines

The Case of Pepper

S. Tosoni, G. Mascheroni, F. Colombo

ABSTRACT

In this chapter, we argue for a sociologically inspired media studies approach to social robotics, and to Human-Machine communication in everyday life contexts. From this perspective, social robots need to be contextualised within the radicalization of the process of convergence, that in the last 30 years has reshaped our media environment. After the convergence of media, in fact, what we are witnessing today is a convergence of media and programmable machines, leading to the rise of media-machines (including, along with social robots, smart speakers and the internet of things). Advanced automation allows these new devices not only to exert some form of material agency in the world, but also to propose themselves to humans as proper communicative partners, interacting with them and co-constructing meaning in a 'natural way'. Our study investigated the relationships between humans and media-machines in natural settings, drawing on two cases of human-robot communication: the deployment of Pepper at the Bologna airport, and at Università Cattolica del Sacro Cuore – curated by the authors. The preliminary results show how the interaction with social robots, far from being 'natural', is actually prefigured and disciplined by the robot programming, configuration and scripts, and how humans are instructed to assume a defined – and limited – interactional role.

Introduction

Social Robots are (partially) automated and algorithmic-based devices designed to engage in "quasi-communication with human beings" (Hepp, 2019, p. 78): they mimic "the affective dynamics of human relationships" (Zhao, 2006, p. 402) to enter in a communicative relationship with their users (Breazeal, 2004). Contrary to other semi-automated media (such as algorithmic bots or digital assistants), social robots' quasi-communication is embodied: by virtue of its material presence in the environment, the device directly participates in a multisensorial interaction with its users – for example, through sound, voice, movement, touch. In the case of humanoid social robots, this presence in space also entails real-time voice interaction, proxemics, body posture, touch and

complex sequences of movements. This ability to simulate embodied human interaction is granted by the centrality of what, in actor-network theory terms, is called the material agency (Knappertt & Malafouris, 2008) of the device, or its ability to alter the course of action of another actor in the material world (Sayes, 2014).

From this point of view, social robots are part of a broader range of new and apparently diverse technological products that in the last ten years have witnessed increasing success on the consumer market, and are populating everyday life contexts to a greater and greater degree, including: smart toys, smart speakers, the Internet of Things (Bunz & Meikle, 2018), such as smart home appliances (e.g., smart fridges, smart washing machines, robot vacuum cleaners, and the like) and entertainment technologies (smart TVs, smart sound systems, etc.). However diverse in functionalities and intended uses, common to all these devices is the fact that they combine communicative features (as media) with the capability to exert material agency in the world (as machines). Smart speakers like Amazon Echo, or Google Home can – for example – broadcast music, podcasts or audiobooks; allow direct communication between users within and outside the house; adjust the color and the intensity of the lights; turn switches off and on and operate (robotic) vacuum cleaners, coffee machines, and other devices connected to the internet. In this sense, internet-connected things "gain new skills that are expressed in new forms of communication" (Bunz & Meikle, 2018, p. 1): they can track, address, see and speak and, in so doing, reconfigure the agency of both material objects and users. Moreover, these actions can be concatenated to form complex consequential or conditional (if/then[1]) chains of action. We will refer to this class of devices as media-machines (Colombo, 2020).

So far, social robots and media-machines (Henschel et al., 2021) have been the object of the vast literature on the multidisciplinary field of human-machine interaction (HMI), aiming primarily at designing, optimizing and evaluating their functioning, their interfaces and their forms of quasi-communication. More recently, the emerging field of human-machine communication (HMC) proposed to conceive of human-robot interactions as a form of communication, that is, as a co-construction of meaning between human actors and technology (Guzman, 2018). From the standpoint of HMC, technology does more than function as a transmission channel: it is re-positioned as a "communicative subject, and it is this subjectivity, rather than interactivity, that marks this technological

[1] IFTTT (If This Then That) is the name of the most diffused software to integrate the functioning of different smart devices into a conditional chain of action. See https://ifttt.com/.

transition" (Guzman, 2018, p. 17). What is at stake here is a major shift in emphasis: HMI, in fact, addressed the anthropomorphic features that are embedded in the materiality of the technology and in its programming, with the intent of improving the design and the behavior of social robots in order to make their human-like nature more consistent and credible to users; HMC, instead, brings to the fore the ways social robots as media-machines come with a social role inscribed in their design and programming – usually that of an assistant – and the ways this role is later enacted in the interactions with humans (Guzman, 2019). In other words, the focus is now on the situated practices of engagement with social robots, and the meanings that emerge from such interactions. This theoretical shift is accompanied by a methodological shift, from research in experimental settings to research in the everyday-life contexts in which social robots as media-machines are employed and used.

However, notwithstanding the emergence of HMC as an area of research, a sociologically inspired and media-studies take on social robots as media-machines is still at its early stage. In what follows, we make the argument that we have much to gain from an approach of this sort if we aim to understand the role of media-machines in reshaping the practices and routines that constitute the fabric of our daily lives. To do so, we will proceed in three steps. First, we will contextualize media-machines and social robots as objects of study within the media-studies tradition, interpreting their advent as the radicalization of the process of convergence that has been reshaping the mediascape since its full digitalization. Second, we will clarify the contribution that the tradition of media-studies research – particularly when more sociologically inspired – can make to the understanding of the phenomenon: after reviewing current literature on the topic, we will advocate for a reframing of the domestication approach to tackle our relationships with media-machines. In the last section, we will draw on the preliminary findings of two ongoing research projects on the deployment of Pepper – one at Bologna Guglielmo Marconi Airport; the other, still in an exploratory phase, at Università Cattolica del Sacro Cuore of Milan (UCSC) – to outline the main tenets of an approach of this sort applied to social robots.

The Advent of Media-Machines

Media-machines are making their appearance at the verge of two distinct and yet interrelated processes of transformation affecting our media and technological landscapes. The first is what can be called the *botization of media*, consisting of the increasing automation of their functioning (Chan-Olmsted, 2019). Early scholars of digital media had al-

ready acknowledged automation as one of the most innovative features of the then new media. For example, Lev Manovich observed how the "numerical coding of media […] and modular structure of a media object […] allow [us] to automate many operations involved in media creation, manipulation and access" (2001, p. 53). For Manovich (2001), the automation of access, in particular, "led to the next stage in media evolution: the need for new technologies to store, organize and efficiently access […] media materials" (2001, p. 55). Twenty years later, the algorithmic re-intermediation of contents represents one of the key assets of the media infrastructure of the so-called 'platform society' (Van Dijck et al., 2018) and the linchpin of the new attention economy (Goldhaber, 1997) as its main model of valorization. When compared to 'traditional' video repositories and pay TVs, for example, recent video platforms like Netflix (Jenner, 2019) introduce a new kind of interactivity with their users. While sharing with previous media the same responsiveness, immediacy and reactivity to users' inputs, these platforms actually gather and algorithmically process data on viewers' behaviors and preferences, constructing complex classifications of their audiences, and autonomously channeling them into patterns of consumption through suggestions and recommendations (Napoli, 2010; Webster, 2016). A similar, often opaque, automated form of content mediation is performed by social-media platforms or by search engines: from this point of view, our interaction with these platforms is closer to the interaction we have with autonomous machines than it is to our interaction with 'old' analogue media.

Simultaneously, we observe an ongoing process of *mediatization* of our machines – that is, the increasing array of communicative functions attributed to commonly used technological tools that, in our daily lives, mediate, enable or extend our agency in the world. This process follows and relaunches the advent of what Mark Weiser, back in 1988, envisioned as *ubiquitous computing* (Greenfield, 2010), where "'ubiquitous' meant not merely 'in every place,' but also 'in everything'. Ordinary objects, from coffee cups to raincoats to the paint on the walls, would be considered as sites for sensing and processing of information, and would wind up endowed with surprising new proprieties" (Greenfield, 2010, p. 11). This paradigm, that Adam Greenfield (2010) termed *everyware*, allowed these electronic devices to work in an increasing automated and programmable way. Moreover, it allowed the interoperability of these machines: thanks to data sharing, they are in fact becoming integrated in complex interlaced ecosystems that enable the coordination of their activities. As a consequence, these machines are increasingly endowed with communicative functions – between one another, and, through programmable interfaces of different sorts, with their users: they report

on their status, update them on their activities and inform their users on the execution of their program. Notably, the infrastructures and organizational networks allowing their interlaced operation and their communicative functions is commonly the same that supports our pervasive media ecologies: the internet. In the Internet of Things (Bunz & Meikle, 2018; Greenfield, 2018), media and machines share the same informational environment, the same codes and, increasingly, the same operating protocols. Yet, the Internet of Things is more than connecting objects to the internet and fitting them with sensors: it is about their mediatization. Once embedded with sensors, software and connectivity that support the exchange of data, machines are turned into media that "mediate what has not been mediated before" (Buntz & Meikle, 2018, p. 18), thanks to their newly acquired ability to gather, generate and distribute information on their users and the surrounding environment – they can track, address, see and speak (Bunz & Meikle, 2018).

Today, the interrelation between the bot-ization of media and the mediatization of machines is paving the way for a radicalization of the well-studied process of media convergence (De Sola Pool, 1983). Media convergence, in fact, has blurred the distinction between media devices that were once specialized as a single mode of communication: a single medium like a smartphone, for example, can be used to make a phone call, play videogames, watch a television series, listen to music and browse the internet, as a wayfinding technology and so on and so forth. The convergence of media and machines, as we have seen, is populating our daily lives with new hybrid devices that integrate a plurality of modes of communication with the capability to exert different forms of material agency, like social robots. The advent of media-machines does not only reconfigure the status and agency of machines, it also challenges consolidated notions of media. In fact, such technologies "are not simply media in the sense that they serve as interaction nodes between people" (Hepp, 2019, p. 79). Rather, as we have already mentioned, media-machines are generally perceived and addressed as communicative partners endowed with their distinctive subjectivity (Guzman, 2018, 2019). We do not simply communicate *through* such media; we now communicate *with* such media. For media scholars, this represents a challenge: to pursue the traditional research questions of their field they are today called not only to update their research agendas to include objects – like, for example, vacuum cleaners (Gross, 2020) – that once fell well behind the border of their disciplinary domain; they are also called to vigorously rethink their methodological and theoretical frameworks. As Peter & Kühne (2018) have argued, first, social robots as media-machines challenge our notions of media, insofar as they are communicative partners. Second, they challenge our understanding of what a communication part-

ner is, since, so far, this subject position has been exclusively attributed to humans. Finally, social robots as media-machines challenge our notions of the boundaries of (human) communication. Therefore, in the next section we will discuss these challenges as they are encountered and addressed by a line of investigation in media studies: the domestication-of-technology approach, interested in analyzing the role played by media(-machines) and social robots in the constitution of our daily lives.

The Domestication of Social Robots as Media-Machines

Like any other media, social robots as media-machines are not transformative *per se*: it is the way they are 'domesticated' and integrated into the media repertoire of individuals and groups, and the way they are made meaningful within socio-material contexts, that actualizes their transformative potential. The process of domestication involves the wide range of meaning-making practices through which technologies are rendered meaningful and useful in everyday life contexts – that is, how technologies are appropriated, adapted as much as adopted, negotiated, even resisted by social actors situated in social contexts (Silverstone et al., 1992; Haddon & Silverstone, 2000). From this perspective, the domestication of media is never fully pre-determined by the functionalities of the technology, nor by its discursive construction in commercial and social representations; rather, it is always contextualized and shaped by the social, cultural and power dimensions that characterize a specific context of use.

In examining the symbolic and agentic practices through which the media are appropriated and made sense of, the domestication of technology approach offers a distinctive and insightful framework for the understanding of the social construction of media technologies and their use in context (Courtois et al., 2012). For example, De Graaf et al. (2018) studied the incorporation of an internet-connected social robot into the domestic environment. The results show that domestication evolves over time as users become more familiar with the technology and the novelty effect fades. A similar conclusion is drawn by Lopatovska et al.'s study (2019) on the use of smart speakers in the domestic context, which suggested that usage declines and varies over time based on the length of ownership and the participants' perceived satisfaction with their interactions with Alexa, resulting in some functionalities being reduced or dismissed. The study has also shown that smart speakers are mainly positioned in the living room or in the kitchen, thus encouraging shared rather than individualized use. In this way, Alexa is domesticated as a family device. Yet, different family members engage in dis-

tinctive activities with Amazon Echo, the most common being: information (including weather forecasts, fact-checking, and news) and entertainment (playing music, telling jokes).

All these studies focus on the objectification, incorporation and conversion of digital media – that is, how digital artefacts are found a place in the space and routines of domestic life, and how their attributed meanings are also converted into symbolic resources for the social construction of users' identities. In this respect, domestication scholars have always emphasized the need to focus on both the material dimension of technologies (their material design; their technical features, the standards adopted, etc.); and their symbolic dimension (i.e., the meanings that are encoded in technology through its commercial and social representations; and the meanings of the media content that is accessed through the technological artefact). This is commonly referred to as the double articulation of information and communication technologies (ICTs) to signal that the media are re-integrated into everyday life as both "objects and media: ICTs are doubly articulated into everyday life as machines and media of information, pleasure, communication" (Haddon & Silverstone, 2000, p. 251). Yet, "doubly-articulated research has proved surprisingly difficult" (Livingstone, 2007, p. 18), resulting in much of the subsequent work in this area prioritizing the symbolic over the material. Most of the work has resolved to focus on the symbolic, that is, on analyzing media as texts or symbolic messages that mediate between the public sphere and the privacy of the domestic sphere. The material has been, at least partly, expunged from the research agenda, which mainly revolved around issues of consumption and representation. However, as the domestication of technology framework was extended beyond the domestic environment (Haddon, 2004), and as the materiality of technological artefacts gained renewed prominence with the variety of portable digital media, scholars have proposed complexifying double articulation theory to better grasp the dimensions in which the domestication of digital media occurs: an update of the original approach that is of pivotal relevance to addressing media-machines and their peculiar forms of material agency. By and large, domestication scholars agree on advocating for at least a third articulation of media – even though the specific proposals about how to conceptualize this third articulation differ. Hartmann (2006), for example, called for an understanding of media as objects, texts and symbolic environments. Drawing on Hartmann, Courtois et al. (2012) propose to focus on the importance of the social and spatial context of media consumption as the third layer mobilized in the meaning-making practices of domestication. Similarly, Lievrouw and Livingstone (2006, p. 23) theorize that digital media comprises three dimensions:

– the artefacts and devices that enable and extend our ability to communicate;

– the communication activities or practices we engage in to develop and use these devices; and,

– the social arrangements or organizations that form around the devices and practices.

Stressing on further aspects, Seija Riddell (2014) talks about representational, presentational, and non-representational aspects that are actualized in media engagement: these three aspects refer respectively to what is represented in the media (representational); to those symbolic resources that are embedded in the media as technological artefacts by their (already mentioned) discursive construction in commercial and social representations (presentational); and to those aspects of our relationship with the media that exceed the symbolic dimension, such as forms of bodily habituation to our devices (non-representational) in particular. This last proposal is in line with recent approaches to out-of-home media domestication, focusing on media engagement in mobility (Moores, 2012) and in spatial contexts that differ from the private space of the house, such as public and semi-public urban space in particular (Tosoni, 2015; Tosoni et al., 2019). Finally, the introduction of (at least) two additional points of attention seems to us to be required specifically by media-machines and social robots, as discussed in the previous section. The first one is related to the communicative role taken up by the social robot, and to the kind of subjectivity actually attributed to it by its communicative partners (Guzman, 2018). The second one refers to the enactment of the media-machine material agency in their contexts of engagement.

All these attempts to re-frame and update the domestication approach led us to outline an analytical framework for the study of social robots as media machines built around three dimensions: 1) the symbolic meanings attributed to the social robot as both a machine and a communicative partner; 2) the material dimension of the media-machine, focusing in particular on its technical features, its embodied presence in space and on the material arrangement of its deployment; 3) the pragmatic dimension represented by the enactment of the robot's agency as a situated interactional performance that involves its users in specific contextual situations, and while engaged in specific practices and routines. It must be stressed, however, that the distinction of these three dimensions is actually purely analytical, as they are inextricably intertwined and activated simultaneously in every instance of interaction with the robot.

We probed this theoretical framework in two cases of deployment of a social robot in public space: a topic so far approached mainly from a HRI perspective (Mubin et al., 2018). With our study, we intend instead

to focus on their outdoor domestication in its triple articulation (symbolic, material and pragmatic).

The (Outdoor) Domestication of Social Robots: The Cases of Pepper at Bologna Airport and at UCSC

The (outdoor) domestication approach to social robots advocated in the previous section was probed and tuned up in an empirical study we began in December 2019, aimed at investigating the ways in which the interaction between humans and social robots is made meaningful within 'natural' socio-material contexts, and the social factors that contribute to shaping the interaction itself. In particular, we aimed to shed light on the ways people's interaction with social robots is molded by the practices and routines they are engaged with, and how these same practices and routines are supported and/or reconfigured by the interaction with social robots. For this reason, we adopted a mixed-method research design to study two deployments of the same model of social robot, Pepper, in a semi-public space: at Bologna Airport and at UCSC – where we are both based – in a deployment we contributed to setting up. In what follows, we will propose some preliminary methodological and theoretical observations on the two cases. Yet, both case studies are to be considered incomplete and still ongoing. The COVID-19 pandemic has in fact disrupted our observation fields: at Bologna Airport, the deployment of Pepper was actually suspended in March 2020 – just two months after our observation started – to avoid gatherings of travelers; in our campus, the deployment of Pepper was delayed by more than one year and was only set up in May 2021, when the university was still dramatically under-attended.

This section is divided in two parts: in the first, we will briefly introduce Pepper from Softbank Robotics, currently one of the more widely known models of social robot for deployment in public space; in the second, we will describe in detail the different multimethod approaches we adopted to investigate the outdoor domestication of the social robot in each site. In the next section, we will present our preliminary findings, organized into the three aspects (symbolic, material and pragmatic) that have emerged as critical for an outdoor domestication approach to social robots.

Introducing Pepper and our case studies

Launched in June 2014 by SoftBank Robotics, Pepper is a humanoid social robot designed as a multi-function assistant for a plurality of con-

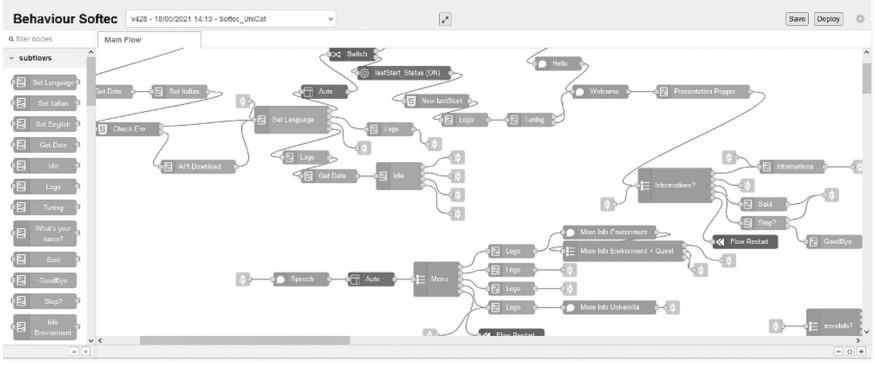
texts of use, and one of the first models to move beyond the prototyping phase and go into mass production – up to June 2021 – for the B2C market (Pandey & Gelin, 2018). Compared to NAO, its predecessor from the same company, Pepper is far taller (1.20 m) and mounted on three wheels that allow it semi-independent movement in space. These features make it more suitable for interaction with humans in public spaces. Moreover, its technical features include actuators and 17 joints that guarantee smooth movements and body gestures; a tablet set on its chest that can be used as an additional interface for communication; a plurality of sensors – including tactile and proximity ones; cameras and microphones that gather data on the surrounding environment to be processed by the robot's algorithms or AI; and a built-in voice synthesizer and speech recognition system. Finally, Pepper is connected to the internet and can therefore dialogue with cloud systems and external services through APIs. As a programmable machine, what Pepper can actually achieve depends on the complexity of its coding, which varies greatly. In the Japanese market, where the use of social robots is more widespread, its limited pre-installed functionalities can be enhanced simply by downloading user-generated behaviors (similar to Amazon's Alexa Skills) that allow the robots to engage its users in games, storytelling or training sessions. The most sophisticated deployments of Pepper, however, are based on the integration of its basic capabilities with more powerful third parties' AI systems, like IBM Watson (Munawar et al., 2018; Varrasi et al., 2019), that can, among the other things, improve its speech or image recognition capabilities, allow the robot to sustain more complex dialogues, or grant it the capacity to move in space in a fully autonomous way.

In both our case studies, however, the robot has been programmed by a middleware, *Orchestra Robotics BSM*, developed by Softec SpA. The robot behaviors that can be programmed with *Orchestra Robotics BSM* are actually very basic (consisting of scripted speeches, gestures and algorithmic procedures triggered by the recognition of keywords). Yet, the middleware has the advantage of making robot programming not only simple, but also understandable by non-coders. In fact, the robot's behavior is represented by a flow-chart (see Figure 1), where each box represents an action: as we will see, the analysis of robot programming, and its commentary by coders, was valuable methodological leverage for tackling its outdoor 'domestication' in both our case studies.

The first of our case studies is represented by the deployment of Pepper at Bologna Airport from the end of March 2018 to March 2020, when it was suspended due to the pandemic crisis. Pepper was located between the check-ins of the airport, a bar, and an information desk, in an area demarcated by mobile barriers and stickers on the ground. This

location was selected after a spatial people-flow analysis within the airport (conducted by an external company) and chosen as an area of intense circulation of incoming and departing passengers. The role assigned to the robot – namely that of information and entertainment – was chosen jointly by the different stakeholders within the airport: its management and the communication, technical and commercial departments. Accordingly, Pepper was programmed to provide passengers with some basic information about the location of several airport services, and at the same time to amuse them with short dances or selfies. At the beginning of 2020 the Airport management was planning to relaunch the project – which was regarded as a communication success due to the level of hype generated in the local and national media – expanding it in two different ways: first, integrating Pepper with the airport's information systems so that the robot could provide passengers with specific information about their flights (e.g., check-in desks, gate numbers, delays); and second, involving the management of in-airport shops among the project stakeholders, to imagine new possible roles for the robot – or for new units to be deployed.

Figure 1 - *A view of Pepper's behavior in the UCSC deployment, as programmed by Orchestra Robotics BMS (Softec SpA)*



The second case is represented by an exploratory deployment we curated in UCSC's Department of Communication. Once again, the site chosen for the deployment was a transit location – the hall of the department building, near the entrance and the custodians' desk – circumscribed by mobile barriers and branded with the university logo. We considered our stakeholders to be our department colleagues, the administrative management of the university and building and the custodians themselves. After several consultation sessions, we assigned Pep-

per four different functions: to provide information on the location of classrooms and other facilities (through spoken text and videos sent to the users' mobiles), on the historical building where the department is located, on air pollution (for an experimental project in environmental communication on Pepper's capacity to raise awareness of the issue), and about the university's anti-COVID health measures.

Addressing the (two-step) domestication of social robots: methodology

As a programmable machine, the 'domestication' of Pepper happens in two main steps: in the prefiguration of the interaction by programming the robot and preparing the material set-up of the deployment, and during the interaction itself[2]. We adopted different methodologies to investigate each step in each case study.

In relation to the programming and set up of the robot at Bologna Airport, we conducted seven in-depth qualitative interviews with the various stakeholders involved in the deployment, in order to investigate how the project was defined, its main goals, and the details of its execution. Concomitantly, we interviewed the Softec Account Manager, the lead engineer, and the engineer who programmed the robot behavior. On this occasion, we could examine its code in *Orchestra*, to discuss the reasons behind some technical choices (for example, setting up a particular threshold of confidence in the speech recognition system to mediate between possible missed and mistaken understanding of a triggering keyword, in relation to the soundscape in the deployment zone). The interviewees also reported their observations about travelers' actual interactions. This knowledge also helped us to set up our own deployment, at UC, and to design the robot behavior in *Orchestra,* which was finally implemented by the Softec engineer. Stakeholders' interviews (n=5) helped us to identify possible tasks to assign Pepper, but also to shed light on our stakeholders' prior understanding of social robots. Moreover, in the UCSC case, we could take advantage of the engineer presenting and providing commentary on his code. This was a valuable methodological entry-point to shed light on how both the users and the interaction are prefigured and scripted into the robot behavior (Akritch, 1992).

The methodology adopted to study step 2 involved ethnographic ob-

[2] A preceding step whereby a general prefiguration of the interaction was inscribed (Akrich, 1992) in the technological artefact is represented by its design and prototyping phase, when crucial decisions were taken, including deciding to make Pepper taller than NAO, providing it with a tablet and wheels, and specific sets of sensors. This step exceeds the scope of our case study.

servations of the situated interactions between users and the robot. In both our case studies, we opted for short observation sessions (1 hour) repeated over time, at different times of day and on different days of the week (Bologna Airport n=10; UCSC n=6). In both sites, for privacy reasons we avoided video-recording the interactions and relied on ethnographic notes instead, taken from a convenient (physically removed) position. In Bologna, observation has so far been complemented with interviews of the waiters in the bar who are exposed to the interaction with the robot on a daily basis, and – as already stated – with various airport management employees. At UCSC, we interviewed people we involved in the testing of the robot prior to its official deployment, not only to correct bugs and fine-tune its behavior, but also to gain an insight into their perceptions of the robot before and after their interaction.

Preliminary Findings

Domesticating Pepper in its symbolic dimension

The domestication of Pepper as a symbolic artefact – that is, as a communicative partner – starts well before the situated encounter with the social robot. In fact, domestication scholars talk of the process of imagination that precedes, while informing it, the acquisition and adoption of technological artefacts. It is through imagination that future users engage in meaning-making practices that will shape adoption and use: they imagine themselves as potential users; they are socialized with new media through their social networks (especially through the experience of early, often enthusiastic, adopters); and extend meanings to the new technology that were previously attributed to other, now familiar, devices; furthermore, they borrow from wider technological experiences such as those propagated in media representation and popular culture. The latter two aspects – namely, the extension of meanings mobilized in the domestication of other digital media to Pepper and the influence of the popular technological ideas about humanoid robots – were particularly relevant in our case studies.

People draw upon prior experiences with media already domesticated in their complex media repertoire in order to reduce the novelty effect. Indeed, designers encode new technologies with 'bridges of familiarity' (Fidler, 1997) to provide links with other media and so facilitate adoption. Yet, 'bridges of familiarity' can equally hinder the process of domestication by creating expectations that the new technology cannot meet, as we could observe in our ethnographic sessions and our own testing of the robot (see section 'The pragmatic dimension of Pepper's domestication').

In fact, users interacting with Pepper are likely to have already domesticated, and therefore familiarized, conversational agents and smart speakers. However, Pepper's conversational agency is less sophisticated and basic in comparison with most conversational agents on the market, first and foremost Alexa. When a user approaches Pepper drawing on their experiences with Alexa, they expect Pepper to understand and reply in real time to virtually any kind of request. Yet, Pepper's conversation is scripted. As a consequence, users' expectations are likely to be frustrated. To avoid what are interpreted as communicative failures due to misplaced expectations on part of the user, the programming phase includes a sort of 'predomestication' by the user and a prefiguration of the interaction. On one side, the user approaching Pepper for the first time is trained by means of a step-by-step introduction that guides them into the social robot's abilities. More specifically, Pepper asks its communicative partners to focus on the color of its eyes and to ask questions only when its eyes are blue, or to repeat their questions when it fails to understand any of the trigger words included in the script. On the other side, the prefiguration of the situated interaction involves several test sessions through which researchers engage with the machine in order to fine-tune the script. A notable example of this process is represented by the need to anticipate all the possible keywords that a user may include in their replies to Pepper, in order to simulate the robot's understanding of the user's needs and requests in nearly real-time and a more effective manner. The several informal interviews with the engineers further illuminated how the conversational agency of social robots such as Pepper needs a careful pre-planning: the conversation with humans must be foreseen and scripted in the machine's code, together with communicative repair strategies in case of failure.

Moving on to the technological collective imagination around social robots, we could observe the influence of both utopian and dystopian discourses on the social consequences of social robots, triggered by the humanoid, embodied form of Pepper. This anthropomorphic form contributes to shape users' expectations, suggesting an emphatic and 'natural' interaction. Moreover, embodied humanoid social robots trigger users on a sympathetic level, which is anticipated in the design and programming phase by engineering entertainment features, such as jokes or dance.

Dystopian imagery associated with social robots was also observed, notably among UCSC custodians, who were initially suspicious, fearing Pepper was being introduced to ultimately automate their labor. As we will elaborate below, the observation of the UCSC building doormen in their interaction with Pepper suggests that, while technological ideas are powerful in shaping initial expectations and interactions with a new technological artefact, as the domestication process continued, the

role of technological imagery was complemented or replaced altogether by direct experience gained through repeated interaction with the machine. This finding is also consistent with current re-formulations of the domestication approach that insist on the need to extend the timeframe of empirical studies and observe the longer-term appropriation of technologies (De Graaf et al., 2018).

Domesticating Pepper in its material dimension

As explained in the theoretical framework above, part of social robots' agency as media machines relates to their material presence in space. Both at Bologna Airport and the hall of the UCSC building, Pepper was located at a crossing point: respectively, on the way from the entrance to the check-in desks, and beside the elevator up to the department and classrooms levels. Pepper's material presence re-structured such spaces. More specifically, in the airport, a space previously intended for the flow of passengers from the entrance to the check-in desks and the gate, was reconfigured to host the robot, behind barriers – to protect it from unintentional impacts – and signs and stickers that tell users where best to position themselves to interact with the robot. In the next section we will discuss the implication of this new material arrangement of space for the practices and routines it hosts: here it is enough to underline that each deployment does not merely involve the robot, but also a wide range of material artefacts that set the stage to host – and try to prefigure – future interaction.

Yet, it's not only space that must be adapted to the presence of the robot: the robot itself must be adapted to the specific physical features of the space that will host it. Both at the airport and in USCS, in fact, the interaction is initiated by Pepper calling for attention and inviting passengers to stop and engage with it. This is a crucial aspect of Pepper's programming: more specifically, not only does the script need to include Pepper's invitations to engage and the careful scheduling of the intervals between one invitation and the next; but the programming also needs to involve aspects such as setting the volume of Pepper's voice, minimizing its' ability to capture and respond to surrounding noise, the physical distance at which Pepper starts engaging with its interlocutors (usually, it is set up to interact with users located one meter in front of it). Expected users must also be taken into consideration when programming the robot: at the airport, in particular, Pepper also had to be set up to recognize the voices of children. The airport staff we interviewed explained that their own children were invited during the programming phase to expand the range of voice tones that Pepper is able

to understand and respond to. This was not needed at UCSC, where the presence of children is not expected.

Finally, one sensitive aspect related to Pepper's material presence in space has to do with its humanoid form, which invites users, children primarily, to engage in haptic, tactile interaction. In fact, the robot was broken after users tried to hold its hand, and a finger went missing. Subsequently, the pre-domestication of users in the script was refined to include explicit invitations to avoid touching the robot.

The pragmatic dimension of Pepper's domestication

All domestication scholars agree that the forms of people's engagement with media depend on the practices and routines they are involved in. As a consequence, these same practices and routines play a relevant role in defining different paths and outcomes of media domestication. Shaun Moores (2012), for example, underlines how, in an airport, media – such as ambient media – are domesticated differently by travelers and workers, since the latter – involved in working practices – have different, recurring and longer forms of exposure to and engagement with them. This is actually what we observed both at Bologna Airport and in UCSC, where workers and travelers (Bologna), and doormen and students (UCSC) engaged with and domesticated Pepper in different ways.

In both sites, the robot was placed in transit points. Regarding travelers and students, this meant that Pepper was encountered while on the move, walking to enter/exit the gates, or to enter/exit the university building. This is the reason why the robot, as already explained, must at first get their attention, inviting them to divert from their usual routine to stop for a while. Whatever the functions implemented in the robot behavior are, our observations indicated that the main motivation for this diversion was the technological awe that authors like Nye (1996) or Mosco (2005) consider to be typical of our relationship with radically new technologies at the early stage of their dissemination. This awe makes interaction with the robot a goal in itself, with no particular purpose other than enjoying the surprising novelty of the technical features of the media-machine. The services offered by Pepper, both at the airport and at UCSC, are accessed not because they are needed, but as a way of discovering – and playing with – the robot: for example, students were asking for the location of a classroom just before exiting the university building.

After accepting the robot's invitation to stop and interact, the practice of walking and traversing, typical of the space of deployment, seems to highly relevant in influencing forms of interaction with the machine.

In fact, sessions are usually short and performed in a hurry. Even when travelers or students take more time to interact with the robot, this happens under the gaze of other students/travelers, often waiting their turn and somehow – probably unintentionally – hurrying them. The interaction sessions may happen alone or in groups (with parents usually introducing and mediating the media machine to their children): in this latter case the sessions are longer, with the group taking turns, giving suggestions to the user directly engaged in the interaction (Pepper engages with one user at time) and commenting on the robot's behavior. In any case, interaction time is always just a few minutes. This makes it difficult for users to fully explore the robot's behavior and get really acquainted with it: this contributes to hiding the limits in the robot's programming and keeping the initial sense of awe alive.

As the practice of walking and traversing help mold the interaction with the robot, the deployment of the robot changes the practices that take place on the deployment sites, and in particular how people walk through space. In the previous section, we already described the spatial arrangement of the deployment. In both sites, we could observe how the space became structured into three different symbolic areas, each of them with its own pragmatic rules, particularly concerning access: first, the portion of space where the flow of passengers (rushing to their gates) or students (entering or exiting the university building) remains unaltered by the robot's presence. We labelled this portion of space the *area of non-engagement*. The closest area to the robot (from one to two meters) is the *area of engagement*: this area is accessed by individual users or small groups. Different dynamics emerge among users in the same groups. For example, at the airport, initial access to Pepper by children is usually mediated by parents. In any case, strangers very rarely enter the area if the robot is already engaged. Most often, they wait their turn in a third and *intermediate area* (two to five meters from the robot) together with other people who do not directly interact with Pepper but have stopped to observe other people engaging with the social robot. Notably, unlike the engagement area, the intermediate area is often shared by strangers.

The rules regulating the use of space do not apply to workers, who domesticate the robot differently. In particular at UCSC, in fact, doormen freely access the engagement area to give spontaneous and unrequested support to students during the interaction. They explain to awed students what the robot is able to understand and perform and what it is not; they teach them how and when to speak to Pepper (slowly and loudly, when its eyes turn blue); they suggest the correct keywords to trigger the robot's scripts. In sum, doormen instruct users, and in this way they 'help' Pepper to exert its agency more smoothly, forming with

it what – in Goffman's terms (1959) – could be defined a spontaneous interactional team.

Here two points are particularly relevant: first, the students' and doormen's evaluation criteria for interaction with the robot differ, as proof of different understanding and ultimately domestication of Pepper. As anticipated in the section 'Domesticating Pepper in its symbolic dimension', in fact, students' expectations are defined by their experience with other media-machines (primarily, digital assistants) and are often frustrated by the robot's behavior: when the robot 'fails' to match their expectation, they consider that a technical failure of the media-machine has occurred. The doormen, on the contrary, have learnt to judge the performance of the robot in different terms, thanks to a longer period of interaction and to their collaboration with researchers (including us). When 'unfulfilled expectations' occur, they apply what they have learnt about the functioning of the robot to discern whether to 'put the blame' on technical problems or flaws in the robot's scripts, or – conversely – on 'mistakes' made by the users (for example, when students talk to the robot too fast or when it is not in listening mode). Second, they spontaneously 'teach' this different criteria of evaluation of the robot's performativity to students, clarifying when they are not interacting 'correctly', or when it is the robot that is not behaving as expected. At the Bologna Airport, we did not observe workers adopt such a spontaneous, active role, even if we did observe them – as was also noted in the interviews – acting as 'ambassadors' of Pepper, and tutors for their friends or relatives occasionally travelling by plane or accessing the airport just to play with the robot.

Final Remarks

In our chapter we have advocated for a domestication approach to social robots that considers them to be examples of media-machines emerging on the cusp of the radicalization of the process of convergence, whereby the distinction between media (as communication devices) and machines (as tools materially acting in world) is becoming blurred. To outline the main tenets of such an approach, we reviewed the recent calls to revise the domestication approach and synthesized them in a methodological framework built around three main pillars: symbolic, material, and pragmatic. We then explored this approach with two case studies of the deployment of Pepper. These preliminary findings need to be corroborated with further research in order to shed light on the complexity of the process of domestication on both observation sites. For example, and in particular, we have highlighted how interaction with the ro-

bot is molded by the technological awe inspired by its novelty. It would therefore be of great relevance to extend the temporal framework of the UCSC fieldwork project in order to understand how the domestication of robots changes over time, when students start to become acquainted with its presence. This could be compared with what happens at Bologna Airport, where only frequent flyers are likely to become accustomed with a media-machine that is not easily encountered in other public spaces. Nonetheless, we think that the approach outlined has proved apt for demonstrating the multiple dimensions of the domestication approach (occurring in two steps, and involving a symbolic, material and pragmatic dimension), but also how the process of domestication can have different outcomes depending on the kind of user and the practices and routines they are engaging with.

References

Akrich, M. (1992). The de-scription of technical objects. In W.E. Bijker, & L. Law (Eds.), *Shaping Technology/Building Society: Studies in Sociotechnical Change* (pp. 205-225). MIT Press.

Breazeal, C. L. (2004). *Designing Sociable Robots*. MIT Press.

Bunz, M., & Meikle, G. (2018). *The Internet of Things*. Polity Press.

Chan-Olmsted, S. M. (2019). A review of artificial intelligence adoptions in the media industry. *International Journal on Media Management, 21*(3-4), 193-215.

Colombo, F. (2019). *Ecologia dei media. Manifesto per una comunicazione gentile*. Vita & Pensiero.

Courtois, C., Verdegem, P., & De Marez, L. (2013). The triple articulation of media technologies in audiovisual media consumption. *Television & new media, 14*(5), 421-439.

de Graaf, M. M., Ben Allouch, S., & van Dijk, J. A. (2018). A phased framework for long-term user acceptance of interactive technology in domestic environments. *New Media & Society, 20*(7), 2582-2603.

De Sola Pool, I. (1983). *Technologies of Freedom*. Belknap Press.

Fidler, R. F. (1997). *Mediamorphosis: Understanding New Media*. Pine Forge Press.

Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Doubleday.

Goldhaber, M. H. (1997). The attention economy and the net. *First Monday, 2*(4). https://firstmonday.org/article/view/519/440.

Greenfield, A. (2010). *Everyware: The Dawning Age of Ubiquitous Computing*. New Riders.

Greenfield, A. (2018). *Radical Technologies: The Design of Everyday Life*. Verso.

Gross, J. (2020). Interviewing Roomba: A posthuman study of humans and robot vacuum cleaners. *Explorations in Media Ecology, 19*(3), 285-297.

Guzman, A. L. (2018). What is human-machine communication, anyway?. In L. Guzman (Ed.), *Human-Machine Communication: Rethinking Communication, Technology, and Ourselves* (pp. 1-28). Peter Lang.

Guzman, A. L. (2019). Voices in and of the machine: Source orientation toward mobile virtual assistants. *Computers in Human Behavior, 90*, 343-350.

Haddon, L., (2004). *Information and Communication Technologies in Everyday Life*. Berg.

Haddon, L., & Silverstone, R. (2000). Information and communication technologies and everyday life: Individual and social dimensions. In K. Ducatel, J. Webster, & W. Herrman (Eds.), *The Information Society in Europe: Work and Life in an Age of Globalization* (pp. 233-258). Lanham, Rowman and Littlefield.

Hartmann, M. (2006), The triple articulation of ICTs: Media as technological objects, symbolic environments and individual texts. In T. Berker, M. Hartmann, Y. Punie, & K. Ward (Eds.), *Domestication of Media and Technology* (pp. 80-102). Open University Press.

Henschel, A., Laban, G., & Cross, E. S. (2021). What makes a robot social? A review of social robots from science fiction to a home or hospital near you. *Current Robotics Reports, 2*(1), 9-19.

Hepp, A. (2019). *Deep Mediatization*. Routledge.

Jenner, M. (2019). *Netflix and the Re-Invention of Television*. Palgrave.

Knappett, C., & Malafouris, L. (2008). *Material Agency: Towards a Non-Anthropocentric Approach*. Springer.

Lievrouw, L. A., & Livingstone, S. (2006). Introduction to the first edition (2002): The social shaping and consequences of ICTs. In L. A. Lievrouw, & S. Livingstone (Eds.), *Handbook of New Media* (pp. 15-32). Sage.

Livingstone, S. (2007). On the material and the symbolic: Silverstone's double articulation of research traditions in new media studies. *New media & society, 9*(1), 16-24.

Lopatovska, I., Rink, K., Knight, I., Raines, K., Cosenza, K., Williams, H., Sorsche, P., Hirsch, D., Li, Q., & Martinez, A. (2019). Talk to me: Exploring user interactions with the Amazon Alexa. *Journal of Librarianship and Information Science, 51*(4), 984-997.

Manovich, L. (2001). *The Language of New Media*. MIT Press.

Mascheroni, G., & Holloway, D. (2019). Introducing the internet of toys. In G. Mascheroni, & D. Holloway (Eds.), *The Internet of Toys: Practices, Affordances and the Political Economy of Children's Smart Play* (pp. 1-22). Palgrave.

Moores, S. (2012). *Media, Place and Mobility*. Palgrave Macmillan.

Mosco, V. (2005). *The Digital Sublime: Myth, Power, and Cyberspace*. MIT Press.

Mubin, O., Ahmad, M. I., Kaur, S., Shi, W., & Khan, A. (2018). Social robots in pub-

lic spaces: A meta-review. In S. Ge et al. (Eds.), *Lecture Notes in Computer Science: vol 11357, International Conference on Social Robotics* (pp. 213-220). Springer.

Munawar, A., De Magistris, G., Pham, T. H., Kimura, D., Tatsubori, M., Moriyama, T., Tachibana, R., & Booch, G. (2018). Maestrob: A robotics framework for integrated orchestration of low-level control and high-level reasoning. In *2018 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 527-534). IEEE.

Napoli, P. M. (2010). *Audience Evolution - New Technologies and the Transformation of Media Audiences*. Columbia University Press.

Nye, D. E. (1996). *American Technological Sublime*. MIT Press.

Pandey, A. K., & Gelin, R. (2018). A mass-produced sociable humanoid robot: Pepper: The first machine of its kind. *IEEE Robotics & Automation Magazine, 25*(3), 40-48.

Peter, J., & Kühne, R. (2018). The new frontier in communication research: Why we should study social robots. *Media and Communication, 6*(3), 73-76.

Ridell, S. (2014). Exploring audience activities and their power-relatedness in the digitalised city: Diversity and routinisation of people's media relations in the triply articulated urban space. In F. Zeller, C. Pont, & B. O'Neill (Eds.), *Revitalising audiences: Innovations in European Audience Research* (pp. 236-260). Routledge.

Sayes, E. (2014). Actor-Network Theory and methodology: Just what does it mean to say that nonhumans have agency?. *Social studies of science, 44*(1), 134-149.

Silverstone, R., Hirsch, E., & Morley, D. (1992). Information and communication technologies and the moral economy of the household. In R. Silverstone, & E. Hirsch (Eds.), *Consuming Technologies: Media and Information in Domestic Space*, (pp. 13-28). Routledge.

Tosoni, S. (2015). Addressing 'captive audience positions' in urban space. From a Phenomenological to a relational conceptualization of space in urban media studies. *Sociologica, 9*(3), 1-28.

Tosoni, S., Krajina, Z., & Ridell, S. (2019). The mediated city between research fields: An invitation to urban media studies. *International Journal of Communication, 13*(2019), 5257-5267.

van Dijck, J., Poell, T., & Waal, M. (2018). *The Platform Society: Public Values in a Connective World*. Oxford University Press.

Varrasi, S., Lucas, A., Soranzo, A., McNamara, J., & Di Nuovo, A. (2019). IBM cloud services enhance automatic cognitive assessment via human-robot interaction. In G. Carbone, M. Ceccarelli, & D. Pisla (Eds.), *New Trends in Medical and Service Robotics. Mechanisms and Machine Science*. (pp. 169-176). Springer.

Webster, J. G. (2016). *The Marketplace of Attention: How Audiences Take Shape in a Digital Age*. MIT Press.

Zhao, S. (2006). Humanoid social robots as a medium of communication. *New Media & Society, 8*(3), 401-419.

5. Human Perceptions of Robotics in Agriculture

M. Gatti, F. Manzi, C. Di Dio, G. Graffigna, P. Guadagna, A. Marchetti, G. Riva, S. Poni

ABSTRACT

In recent years, the agricultural sector has been witnessing a technological change characterized by the automation of various activities also due to the use of some robotic technologies. As for the field of application, the pioneering solutions available on the market aim at performing non-selective operations. A key point of success in this field of research is the implementation of cognitive human processes in the algorithms driving robot actions. However, the perception that people have of the use of robots in agriculture is still little studied. The current research is a preliminary study aiming at describing human perception of different robotic solutions that are already available on the market and/or that will be potentially developed in future. An online survey analyzed different case studies encompassing several robotic platforms that are expected to perform selective vs. non-selective and high risk vs. low-risk operations in viticulture. The study evaluated human perception of 49 students mainly enrolled at the faculties of education and psychology with a moderate knowledge in agriculture. The results showed that people prefer the use of humans over robots to perform selective agricultural activities, thus possibly reflecting greater reliability associated to humans in making decisions and performing selective tasks. On the other hand, when comparing the different robots, the humanoid robot 'replaces' the human, i.e., people consider it as the robot to be used in selective activities requiring more complex decisions. This is also confirmed when analyzing the perception on the effect that the introduction of humanoid robots may have in different domains of agriculture: there is a preference for the use of humanoid robots for performing selective tasks in both the safety and quality domains. Finally, autonomous vehicles are preferred to humanoid robot in terms of increased work productivity and reduced risk. Generally, these findings allow us to hypothesize that people place human and robots on two different ontological levels with respect to their mental characteristics, as reported by the participants' preference for using robots in tasks that do not require complex decisions. This consideration is further supported by the few differences found between humans and robots in both high and low risk non-selective tasks.

Introduction

The use of robots and people's perceptions of these technologies have been widely studied in different sectors of our society. Think of the wide use of robots in the manufacturing and industrial context (Matheson et al., 2019) where some forms of human-robot collaboration are necessary and currently possible. Robots are also used in other sectors such as healthcare and education, although with greater difficulty than in manufacturing due to the complexity of the human components involved in these interactions (Marchetti et al., 2020).

Given the long tradition of automation in the manufacturing sector, the introduction of robots seems to be an expected step in its development, although the challenges of collaborative robots (cobots, Colgate et al., 1996) are still many and not completely overcome. On the other hand, the service sector, such as healthcare and education, is not easy to characterize with the use of robots to support the work of humans, since the relational components that connote these interactions represent one of the greatest challenges of human-robot interaction; however, robots are used in several projects in healthcare and educational contexts in which collaboration with humans is possible (Baxter et al., 2017; Cavallo et al., 2018; Casas et al., 2020).

The type of robot used in a specific sector depends on the function it has to perform and the working environment. For example, in the manufacturing sector, robots with mechanical features are mostly used, such as mechanical arms used in the automotive industry, which have to interact with adult workers. In contrast, healthcare and educational contexts prefer robots with a more humanoid design as they have to interact with people with physical and/or psychological weaknesses and with sensitive population groups, such as children and the elderly, with whom it is essential to create a relational engagement. In the agricultural sector, the type of robot to be used and the effect it may have on people's acceptance is a field of research that has yet to be explored.

In recent years, the agricultural sector has also witnessed a technological change characterized by the automation of various activities, including the use of some robotic technologies (Sparrow & Howard, 2021). Agriculture and automation are linked by a narrow relationship that is continuously evolving. After the adoption of modern agriculture during the XVIII century, the introduction of mechanical solutions post-World War I, and the green revolution that characterized the second half of the past century, agriculture is now entering its fourth phase that is commonly known as the digital revolution. Indeed, if mechanization allowed the replacing and/or supporting of several manual operations by mechanical tools (Poni et al., 2016), the food industry is undergoing

a fundamental transformation. Digitalization in agriculture is expected to have a positive impact on production systems thanks to the introduction of enabling technologies such as precision farming (including remote sensing and variable rate technologies), the dissemination of Information and Communications Technologies (ICT), and new challenging solutions such as robotics.

As part of this evolution, smart machines equipped with mechatronic components that combine sensors, electronic control and actuators already exist, allowing the automation of simple operations such as real-time volume adjustment during pesticide spraying, variable rate application of fertilizers and seeds, and animal feeding (Gatti et al., 2018; Vougioukas, 2019; Oliveira et al., 2021). As part of this innovation process, robotics will assume increasing importance and new autonomous platforms will soon be available. Agricultural robots may be classified into one of the following categories: autonomous vehicles for the execution of non-selective operations; autonomous or remotely-controlled platforms equipped with specific sensors for crop and soil monitoring; platforms equipped with robotic arms connected to an end effector (e.g. scissors for pruning, grasping systems for harvesting, etc.) that allow highly selective operations.

As far as field applications are concerned, the pioneering solutions available on the market aim to perform non-selective operations such as canopy spraying executed by autonomous tractors, vineyard floor mowing using small robotic devices, or non-destructive assessment of fruit composition performed by autonomous platforms equipped with specific sensors. However, excluding special farming environments such as greenhouse cropping systems and vertical farming where robotics is already a reality, the use of such technologies to drive selective operations at field scale is relatively far away and few prototypes are currently developed by several institutions and private companies (Botteril et al. 2017; Williams et al., 2020; Kootstra et al., 2021). These systems require artificial intelligence algorithms (e.g., deep learning and artificial neural networks) for the development of optical tools that allow the detection of operational scenarios as well as the identification of optimal actions to be performed through mechatronic arms. A key point of success in this workflow is the implementation of the cognitive human process in the algorithms driving robot actions. Robotics is expected to solve the dramatic shortage of skilled labor across different sectors of modern agriculture as well as make work more productive, timely, effective and safe. However, the perceptions that people have of robot use in agriculture is still little studied.

The current study aims at describing human perceptions of different robotic solutions that are already available on the market and/or that will be potentially developed in future. The study was based on an on-

line survey, analyzing different case studies encompassing several robotic platforms that are expected to perform selective vs. non-selective and high-risk vs. low-risk operations in viticulture.

Materials and Methods

Participants

In July 2021, an online survey was carried out involving 49 Italian university students (81.6% female) aged 19-51 years (M = 24.88; standard deviation [SD] = 6.52). Based on their academic program, the participants were clustered as follows: 74.6% of the panel was from the Faculty of Education, 13.2% from the Faculty of Agriculture and the remaining 13.2% from the Faculty of Psychology. One third of the participants (27.7%) included student-workers, while the others were full-time students (51.4%). Furthermore, 8.5% of attendees were married or cohabiting. Finally, 77.3% of the sample claimed to have medium or higher knowledge of robotics while 61.2% claimed medium or higher knowledge of artificial intelligence. The sample was collected online, and the participants were asked to sign an informed consent and processed in accordance with the Declaration of Helsinki.

Procedure, robotic platforms and tasks

– Procedure

Participants were examined online in a single session that lasted about 20 minutes. Initially, participants were asked about their educational background, family status, job, and general knowledge of robotics and artificial intelligence. In addition, an *ad hoc* questionnaire was administered to analyze their general knowledge of agriculture. Subsequently, participants were asked to read a short text in which it was explained that agriculture is going through a process of digitalization and technological innovation within which robotics have, and will take in the future, an increasingly important role, with an increasing availability of agriculture robots on the market. Furthermore, the text explained that in addition to non-selective operations such as soil management (e.g., weed control) and plant protection (e.g., pesticide application), these robotic solutions might also include selective operations (e.g., pruning of grapevines and fruit trees) performed by way of a robotic arm connected to a specific end-effector (e.g., pruning shears). At the end of this introduction to robotics in agriculture, three robotic platforms were

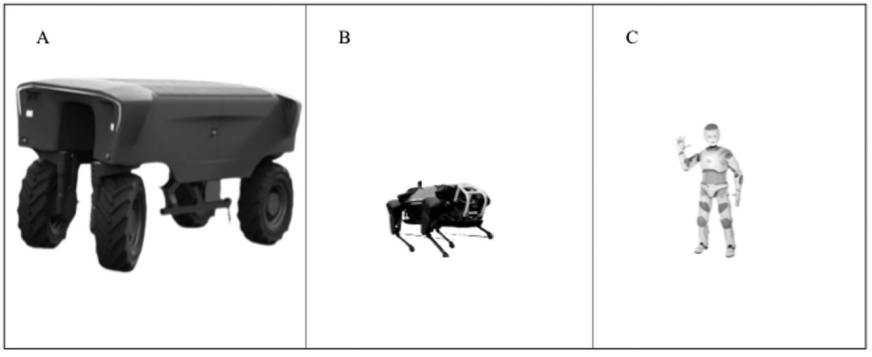
presented with the possibility to be equipped with specific tools to perform selective and non-selective practices. Through this brief introduction, participants were informed about the most recent innovations already available in agriculture as well as the potential solutions due to be introduced by robotics in the mid- to long-term. Participants were further informed of the existence of different robotic platforms (e.g., wheeled, tracked, and legged systems) that, if properly equipped, could be used for agricultural purposes.

The following questionnaires were therefore administered in sequence: Usefulness of Robots in Agricultural Activities, Value of Robots in Agricultural Processes, and Attribution of Mental States.

– Robotic platforms

The three robotic platforms considered as part of the study were: a) the autonomous vehicle 'Bakus' (Vitibot, Reims, France)[1], a universal platform offering a large number of smart and power tools for vineyard management (Figure 1A); b) the quadruped robot HyQReal[2] developed by the Dynamic Legged Systems (DLS) lab of the Italian Institute of Technology (IIT, Genoa, Italy) intended for use in difficult terrains (Figure 1B); c) the humanoid robot Romeo (Softbanks Robotics)[3], a Social Assistive Robot, selected for its extended humanoid characteristics (Figure 1C).

Figure 1A-C - *Representation of the three robotic platforms considered as part of the study: (A) autonomous vehicle, (B) quadruped robot and (C) humanoid robot*



[1] https://vitibot.fr/.

[2] https://www.iit.it/web/dynamic-legged-systems/hyqreal.

[3] https://www.softbankrobotics.com/.

– Tasks

Knowledge on agriculture. Participants' knowledge of agriculture was evaluated through an *ad hoc* questionnaire. This part of the survey was composed of three sections. The first section (general knowledge of agriculture) required the participants to evaluate their own level of agricultural knowledge as based on a 5-point Likert Scale (1 = no knowledge; 5 = professional knowledge). The second section (agricultural processes) was composed of five items that asked participants if they knew (answer options: yes/no) a number of agricultural processes (e.g., agri-food chain, life cycle assessment, water footprint, management of natural resources). The third section (cultural practices) was composed of seven items that questioned participants' knowledge (answer options: yes/no) about some of most important cultural practices (e.g., plant disease and pest control, fertilization, soil management, agricultural machineries). A single score was then calculated, defined as 'Agricultural Knowledge', reflecting the participants' overall knowledge of agriculture. The single score was calculated by summing the scores of the three questions (the score ranges 1-18).

Attribution of Mental States (AMS). The Attribution of Mental States (Manzi et al., 2017, 2020; Di Dio et al., 2019, 2020a,b) is a measure of the mental qualities that participants ascribe when they look at images depicting specific characters (humans or robots). The scale consists of five mental states dimensions (Perceptive, Emotive, Intentions and Desires, Imaginative, Epistemic) and can be calculated in a single score. In the present study, the total AMS score was used for the analysis. This scale is a self-administered tool composed of 26 items evaluated on a 5-point Likert scale (1 = not at all; 5 = absolutely yes).

Usefulness of Robots in Agricultural Activities (URAA). To test participants' opinions on the usefulness of robots in agriculture, an *ad hoc* questionnaire was designed. This was based on two macro areas of agricultural activities, selective and non-selective, which differ in terms of their greater or lower levels of accuracy and decision-making. These activities can be characterized by a different level of worker health risk, which has been assumed as the main factor for identifying high- vs. low-risk operations. Based on the two factors (selectivity and riskiness), the following four combinations were identified and associated with one of most representative operations in viticulture: Selective/Low-Risk (e.g., grape harvesting), Selective/High-Risk (e.g., winter pruning), Non-Selective/Low-Risk (e.g., weed control) and Non-selective/High-Risk (e.g., pesticide application). Participants' opinions on the usefulness of

robotics in performing the above-mentioned activities were assessed. In detail, attendees were asked to evaluate the convenience of performing one of the four operations using one of the selected platforms (Autonomous Vehicle, Quadruped Robot, Humanoid Robot) in comparison to a skilled worker. For the rest of the text, the term *Agents* will be used to refer to the four modalities of operation (the three robotic platforms and the human) when considering also the human, who cannot be included among the robotic platforms. The questionnaire used was a self-administered tool composed of four items evaluated on a 5-point Likert scale (1 = not at all; 5 = absolutely yes).

Value of Robots in Agricultural Processes (VRAP). To test the respondents' opinions on the value of robots in different domains of agriculture, an *ad hoc* questionnaire was set up as part of the present study. The questionnaire was based on four macro areas of agriculture where robots can have an impact compared to human interventions: Safety (i.e., the risk workers are exposed to when performing agricultural activities), Quality (i.e., the accuracy with which agricultural activities are carried in line with requirements), Productivity (i.e., the efficiency with which tasks are completed over time), and Work (i.e., the number of workers and professional figures involved in agriculture). The questionnaire consists of six dimensions and two questions for each dimension: Safety of Selective Activities, Safety of Non-Selective Activities, Quality of Selective Activities, Quality of Non-Selective Activities, Productivity and Work. These dimensions were assessed by considering the three robotic platforms ('How much do you think the following platforms, equipped with specific equipment, can…'). This questionnaire is a self-administered tool assessed on a 5-point Likert scale (1 = not at all; 5 = absolutely yes).

Data Analysis

To investigate the attribution of mental states to the four different agents, the usefulness of robots in agriculture and the value of robots in agricultural processes, an analysis based on three GLM Repeated Measures was carried out. For the GLM analysis, the Greenhouse-Geisser correction was used for violations of Mauchly's Test of Sphericity, $p<.05$. All *post hoc* comparisons were Bonferroni corrected.

Furthermore, to investigate factors possibly associated with the usefulness of robots in agriculture and the value of robots in agricultural processes, Pearson's correlation analyses were carried out between the usefulness of robots in agriculture, and the value of robots in agricultural processes and agricultural knowledge.

Results

Descriptive statistics of agricultural knowledge

The descriptive statistics related to the knowledge of agriculture associated with the 49 online survey participants are reported in Table 1.

Table 1 - *Descriptive Statistics for variables measuring Agricultural Knowledge (General, Processes, and Cultural Practices) (N = 49)*

	Question	*M (SD)*
General	What is your general knowledge on agriculture?	2.33 (.64)
		number of "yes" (%)
Processes	Have you ever heard about food chain?	45 (91.8%)
	Have you ever heard about life cycle assessment?	32 (65.3%)
	Have you ever heard about soil management?	44 (89.8%)
	Have you ever heard about agricultural water footprint?	36 (73.5%)
	Have you ever heard about management of natural resources?	49 (100%)
Cultural Practices	Have you ever heard about plant protection?	46 (93.9%)
	Have you ever heard about crop fertilization?	48 (98%)
	Have you ever heard about soil tillage?	48 (98%)
	Have you ever heard about agricultural machineries?	49 (100%)
	Have you ever heard about crop irrigation?	49 (100%)
	Have you ever heard about planting/seeding?	49 (100%)
	Have you ever heard about crop biodiversity?	49 (100%)

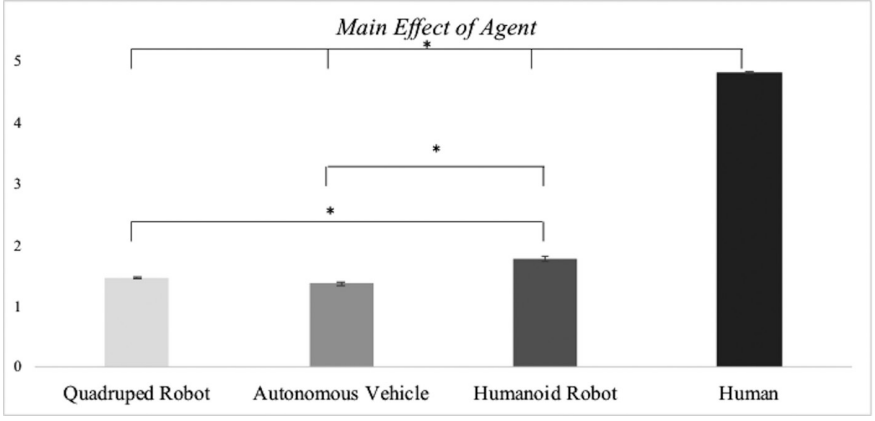
People's evaluations of robot mental qualities

To evaluate the effect of the type of robot on the attribution of mental states (AMS) a GLM Repeated Measures analysis was carried out with four levels of *Anthropomorphism* (Autonomous Vehicle, Quadruped Robot, Humanoid Robot, Human) as within-subject factors. The GLM analysis revealed a main effect of *Anthropomorphism*, $F_{(3,135)} = 823.85$, $p < 0.001$, *partial-*$\eta 2 = 0.95$, $\delta = 1$, indicating a greater attribution of mental states to the Human vs. the robots, and to the Humanoid Robot compared to other robotic platforms (Figure 2). The absolute values asso-

ciated with the participants' scores on the attribution of mental states, identified a significant gap between traditional agriculture characterized by manual operations and robotics (~5 vs. 1.5).

To better discriminate between the Autonomous Vehicle and the Quadruped Robot, a comparison was carried out between the two platforms on all the AMS dimensions. This comparison was conducted to assess whether one of the two outscored on the mental dimensions. For this purpose, we carried out a GLM Repeated Measures analysis with 5 levels of *AMS* (Perceptive, Emotive, Intentions and Desires, Imaginative, Epistemic) and two levels of *Agent* (Autonomous Vehicle, Quadruped Robot). The results showed a significant interaction, $F_{(4,180)} = 32.1$, $p < 0.001$, *partial-*$\eta 2 = 0.42$, $\delta = 1$, indicating that the Autonomous Vehicle scored lower on the Perceptive dimension (Autonomous Vehicle: $M = 1.04$, SE= .02; Quadruped Robot: $M = 1.59$, SE= .11; *M*Diff = .551, SE = .097), but outscored the Quadruped Robot on both the Imaginative (Autonomous Vehicle: $M = 1.08$, SE = .04; Quadruped Robot: $M = 1.17$, SE= .07; *M*Diff = .091, SE = .044) and Epistemic (Autonomous Vehicle: $M = 2.13$, SE = .16; Quadruped Robot: $M = 1.1$, SE = .04; *M*Diff = 1.03 SE = .143) dimensions. From a mentalistic point of view the Autonomous Vehicle is, therefore, considered more anthropomorphic than the Quadruped Robot. Given these findings, the agents considered for the study can be ordered according to their level of anthropomorphization in the following ascending order: Quadruped Robot, Autonomous Vehicle, Humanoid Robot, Human.

Figure 2 - *Participants' scores on the Attribution of Mental States (AMS)*



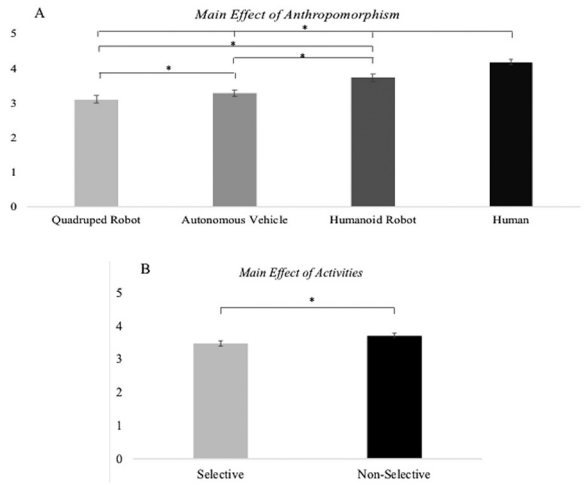
Note: AMS mean scores for the Quadruped Robot (light grey bar), Autonomous Vehicle (grey bar), Humanoid Robot (dark grey bar) and Human (black bar). * Indicates significant differences between two agents.

People's opinions on the usefulness of robots in agriculture

To evaluate the effect of an agent's anthropomorphization on the propensity to use robotic platforms to support agricultural activities, a GLM analysis was carried out with four levels of *Anthropomorphism* (Quadruped Robot, Autonomous Vehicle, Humanoid Robot, Human), two levels of *Activities* (Non-Selective, Selective), and two levels of *Riskiness* (Low-Risk, High-Risk) as within-subject factors. The GLM revealed a main effect of *Anthropomorphism*, $F_{(3,144)} = 31.36$, $p < 0.001$, *partial-*$\eta 2 = 0.395$, $\delta = 1$, indicating that the Human scored higher than all robotic platforms (Human > Autonomous Vehicle, *M*Diff = 0.893, *SE* = 0.114, $p<0.01$; Human > Quadruped Robot, *M*Diff = 1.071, *SE* = 0.133, $p<0.001$; Human > Humanoid Robot, *M*Diff = 0.444, *SE* = 0.114, $p<0.001$; Figure 3A); the Humanoid Robot scored higher than the other robotic agents (Humanoid Robot > Quadruped Robot, *M*Diff = 0.628, SE = 0.137, $p<0.001$; Humanoid Robot > Autonomous Vehicle, *M*Diff = 0.449, SE = 0.143, $p = 0.017$; Figure 3A); and the Autonomous Vehicle scored higher than the Quadruped Robot (Autonomous Vehicle > Quadruped Robot, *M*Diff = 0.179, *SE* = 0.06, $p<0.01$; Figure 3A). Furthermore, the analysis highlighted a main effect of *Activities,* $F_{(1,48)} = 16.54$, $p < 0.001$, *partial-*$\eta 2 = 0.256$, $\delta = .98$, showing that participants are more inclined to use the agents in non-selective operations than selective activities (Non-Selective > Selective, *M*Diff = 0.235, *SE* = 0.058, $p<0.001$; Figure 3B). The main effect of *Anthropomorphism* highlights the scalar impact of anthropomorphization on usefulness in the use of these agents in agricultural activities.

Furthermore, the analysis showed a two-way interaction between *Anthropomorphism* and *Activities*, $F_{(3,144)} = 38.46$, $p < 0.001$, *partial-*$\eta 2 = 0.445$, $\delta = 1$, and a significant three-way interaction between *Anthropomorphism*, *Activities* and *Riskiness* $F_{(3,144)} = 3.24$, $p < 0.001$, *partial-*$\eta 2 = 0.063$, $\delta = .74$. The *post hoc* analyses of the three-way interaction showed that for Non-Selective/High-Risk activities the Autonomous Vehicle is more trusted than the Quadruped Robot (Autonomous Vehicle > Quadruped Robot, *M*Diff = 0.367, SE = 0.104, $p<0.001$; Figure 4A). Also, the *post hoc* analyses revealed that for Selective activities, either High-Risk or Low-Risk, the Human is considered the most entrusting of all the agents (Human > Autonomous Vehicle, *M*Diff = 1.673, *SE* = 0.192, $p<0.001$; Human > Quadruped Robot, *M*Diff = 1.633, *SE* = 0.188, $p<0.001$; Human > Humanoid Robot, *M*Diff = 0.653, *SE* = 0.156, $p<0.01$; Figure 4A), followed by the Humanoid Robot, which is considered more entrusting than the other robots (Humanoid Robot > Autonomous Vehicle, *M*Diff = 1.102, *SE* = 0.203, $p<0.001$; Humanoid Robot > Quadruped Robot >, *M*Diff = 0.98, SE = 0.188, $p<0.001$; Figure 4B).

Figure 3A-B - *Participants' scores on the Usefulness of Robots in Agricultural Activities (URAA)*

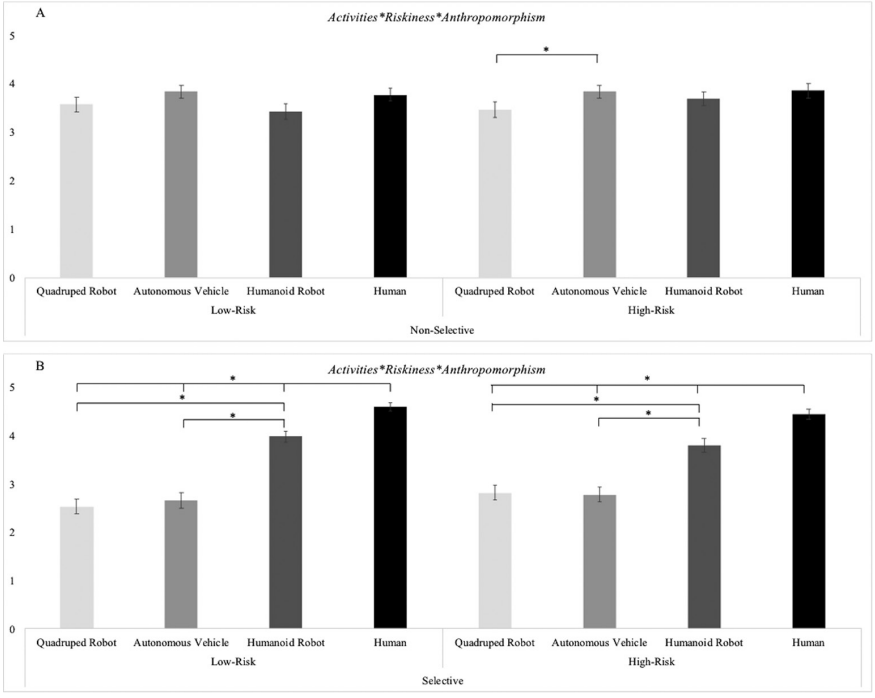


Note: (A) URAA mean scores as a function of Anthropomorphism: Quadruped Robot (light grey bar), Autonomous Vehicle (grey bar), Humanoid Robot (dark grey bar) and Human (black bar); (B) URAA mean scores as a function of Activities: Selective (light grey bar) and Non-Selective (black bar). *Indicates significant differences between two agents (Graph A) and type of activities (Graph B).

People's opinions on the value of robots in agricultural processes

To evaluate people's opinions on the value of robots in different agricultural processes, a GLM analysis was carried out with three levels of *Agents* (Quadruped Robot, Autonomous Vehicle, Humanoid Robot), six levels of *Domains* (Safety of Selective Activities, Safety of Non-Selective Activities, Quality of Selective Activities, Quality of Non-Selective Activities, Productivity and Work) as within-subject factors. For the purpose of this study, we focused only on the significant two-way interaction between *Agents*Domains*, $F_{(3,144)} = 10.38$, $p < 0.001$, *partial-*$\eta 2 = 0.178$, $\delta = 1$, which indicated that the Humanoid Robot scored higher than other robotic agents in Safety and Quality of Selective Activities (Safety of Selective Activities: Humanoid Robot > Quadruped Robot, *M*Diff = 0.653, *SE* = 0.159, $p<0.001$; Humanoid Robot > Autonomous Vehicle, *M*Diff = 0.755, *SE* = 0.176, $p<0.001$; Quality of Selective Activities: Humanoid Robot > Quadruped Robot, *M*Diff = 0.653, *SE* = 0.161, $p<0.01$; Humanoid Robot > Autonomous Vehicle, *M*Diff = 0.673, *SE* = 0.158, $p<0.001$; Figure 5A). Furthermore, the analysis revealed that the Autonomous Vehicle scored

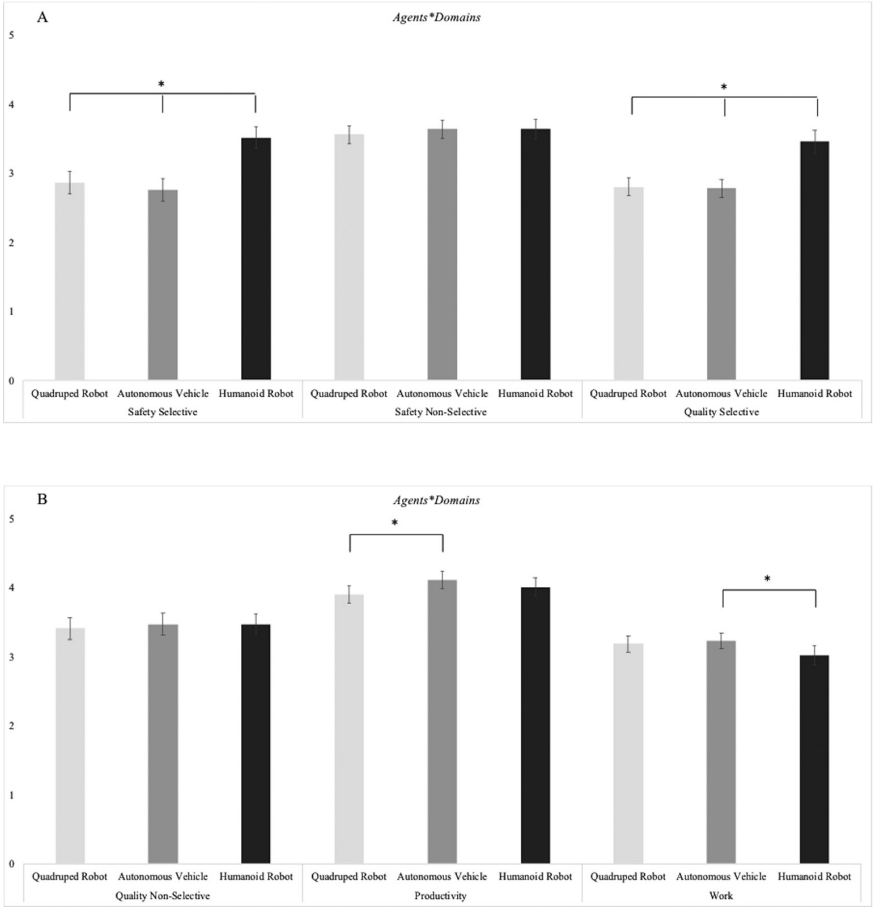
Figure 4A-B - *Participants' scores on the Usefulness of Robots in Agricultural Activities (URAA)*



Note: (A) Graph for the Non-Selective activities, where the URAA mean scores are reported as a function of Riskiness and Anthropomorphism: Quadruped Robot (light grey bar), Autonomous Vehicle (grey bar), Humanoid Robot (dark grey bar) and Human (black bar); (B) Graph for the Selective activities, where the URAA mean scores are reported as a function of Riskiness and Anthropomorphism: Quadruped Robot (light grey bar), Autonomous Vehicle (grey bar), Humanoid Robot (dark grey bar) and Human (black bar). *Indicates significant differences.

higher than the Quadruped Robot in Productivity (Autonomous Vehicle > Quadruped Robot, *M*Diff = 0.102, *SE* = 0.039, $p = 0.033$; Figure 5B) and the Autonomous Vehicle scored higher than the Humanoid Robot in Work (Autonomous Vehicle > Humanoid Robot, *M*Diff = 0.204, *SE* = 0.082, $p = 0.050$; Figure 5B).

Figure 5A-B - *Participants' scores on the Value of Robots in Agricultural Processes (VRAP)*



Note: (A) Graph for the Safety Selective, Safety Non-Selective and Quality Selective where the VRAP mean scores are reported as a function of Agents: Quadruped Robot (light grey bar), Autonomous Vehicle (grey bar), Humanoid Robot (dark grey bar); (B) Graph for the Quality Non-Selective, Productivity and Work where the VRAP mean scores are reported as a function of Agents: Quadruped Robot (light grey bar), Autonomous Vehicle (grey bar), Humanoid Robot (dark grey bar). *Indicates significant differences between two agents.

Correlations

To further investigate the factors potentially associated with the propensity to use robots in agriculture, correlation analyses were carried

out between Usefulness of Robots in Agricultural Activities (i.e., Selective/High-Risk, Selective/Low-Risk, Non-Selective/High-Risk, Non-Selective/Low-Risk) and Agricultural Knowledge (AK). Moreover, to better understand the variables possibly related to people's opinions of the value of robots in different agricultural processes, correlation analyses were carried out between Value of Robots in Agricultural Processes (Safety in Selective Activities, Safety in Non-Selective Activities, Quality in Selective Activities, Quality in Non-Selective Activities, Productivity, Work) and Agricultural Knowledge (AK).

As shown in Table 2, the correlation analysis revealed a positive correlation in Selective/High-Risk activities for the Humanoid Robot. Furthermore, the correlation analysis showed a positive correlation in Non-Selective/High-Risk activities for the Autonomous Vehicle and in Non-Selective/Low-Risk activities for the Humanoid Robot and Human.

As shown in Table 3, the correlation analysis revealed a positive correlation in Safety of Selective Activities and Safety of Non-Selective Activities for the Humanoid Robot.

Discussion

This study examined people's opinions on the usefulness of robots in agricultural activities and people's opinions on the value of robots in agriculture. Specifically, participants' opinions were tested as a function of different robots, which varied with respect to the degree of anthropomorphization.

As expected, the results showed that the participants attributed greater mental qualities to the human than to the robots. Among the three robotic platforms, the humanoid robot was associated with higher mental qualities than the autonomous vehicle and the quadruped robot. However, the autonomous vehicle was unexpectedly evaluated with a more complex mind than the quadruped robot. These results are in line with several studies that have analyzed the effect that different levels of physical anthropomorphization of robots have on the attribution of mental states (Di Dio et al., 2020a,b; Manzi et al., 2020a,b). These, in general, show that the more humanoid the robot, the greater the attribution of mental qualities. The data from the present study extrapolate to other types of robots what was previously found with humanoid robots, depicting a more complex scenario of the attribution of mental states. However, these results are limited to the human perceptions of the specific 49-people panel, which comprised students enrolled in the Faculties of Education and Psychology with only a general knowledge of agriculture, only 13.2% of them having a more solid background. According-

Table 2 - *Overall correlations. Pearson's correlations between usefulness of robots in agricultural activities and the agricultural knowledge of those interviewed*

		ACTIVITIES			
		Selective		*Non-Selective*	
	RISKINESS	*High-Risk*	*Low-Risk*	*High-Risk*	*Low-Risk*
	AGENTS				
AK	*Quadruped Robot*	.253	.271	.236	-.005
	Autonomous Vehicle	.268	.174	**.346***	.099
	Humanoid Robot	**.401****	.186	.131	**.331***
	Human	.264	.096	.224	.411**

Note: *Correlation is significant at the 0.05 level (two-tailed); ** correlation is significant at the 0.01 level (two-tailed).

Table 3 - *Overall correlations. Pearson's correlations between value of robots in agricultural processes and agricultural knowledge*

		AGENTS		
	DOMAINS	*Quadruped Robot*	*Autonomous Vehicle*	*Humanoid Robot*
AK	*Selective Safety*	.194	.165	**.315***
	Non-Selective Safety	.117	.187	**.311***
	Selective Quality	.067	.086	.12
	Non-Selective Quality	.117	.069	-.117
	Productivity	.181	.203	.225
	Work	.174	.174	.211

Note: *Correlation is significant at the 0.05 level (two-tailed); ** correlation is significant at the 0.01 level (two-tailed).

ly, further investigations should be addressed to a wider panel in order to get more balanced proportions among the different participant backgrounds. As a matter of fact, the current survey could be influenced by the fact that humanoids have been largely tested in several disciplines (i.e., education, healthcare, consumer preference, etc.) (Belpaeme et al., 2018; Marchetti et al., 2020) while preliminary agricultural robots are mainly based on unmanned ground vehicles. So, this result requires further studies to be confirmed.

For instance, it would be quite interesting to determine if similar results might be collected in response to the introduction of additional information, such as claiming that, for a given activity, the algorithms driving the identification of the operational scenario and execution of a desired operation were identical irrespective of the robotic platform carrying the operational arm. Coming back to our results, although participants ascribed more perceptive skills to the quadruped robot than the autonomous vehicle (see Figure 1), they attributed more mental ability to the latter. The higher attribution on the epistemic dimension to the autonomous vehicle is particularly interesting, since it is one of the most important factors for building epistemic trust, and it may lead to the hypothesis that people ascribe more trust to an autonomous vehicle than to a quadruped robot. This last suggestion is preliminary and should be investigated in more detail in future studies.

The findings on the effect of anthropomorphization on the attribution of mental states to the agents considered, are also mirrored in people's opinions on using robots in agricultural activities. The data indicated that people would employ the Human more than robots. More specifically, for both high-risk and low-risk Selective activities the human is considered the most reliable agent. This preference for humans over robots reflects the high opinion people ascribe to humans with respect to their knowledge, skills, experience, decision-making abilities and precision characterizing their actions. Indeed, selective activities are characterized by a sophisticated level of decision-making. Let us take grape harvesting as an example: in this activity, the pickers not only perform a quick estimation of fruit properties according to various objective criteria (e.g., bunch health, shape and color), but also decide whether or not to cut a bunch on the basis of processes that are not fully explicable. This complex decision-making process may not be valid for robots regardless of their similarity to humans in terms of physical and mental characteristics. Therefore, it is possible to hypothesize that people place the human and the robot on two different ontological levels, an assumption that is reflected in the usability of robots in certain activities that require less complex decision-making. This statement is confirmed by the scores on the Usefulness of Robots in Agricultural Activities, which showed a reduced gap be-

tween the human and other agents when robotics is associated with both high- and low-risk non-selective operations, such as pesticide application and vineyard floor management respectively. In contrast, the gap between the human and the other agents increases when selective operations such as winter pruning and grape harvesting are considered. Finally, the positive correlation between the involvement of humans in non-selective/low-risk activities and people's agricultural knowledge showed that being familiar with the agricultural work influences people's opinions on the potential use of robots. However, this result also requires further analysis and is strictly limited to the opinions of the respondents involved in the present survey who declared a general knowledge of agriculture equal to 2.3 on a 1-5 scale (Table 1). Indeed, if on the one hand the respondents tended to avoid the association between humans and the execution of high-risk operations, the same panel tended to recommend humans for performing low-risk non-selective operations (i.e., soil management) that could be easily completed by robotic platforms, such as autonomous vehicles equipped with specific tools. In perspective, to better understand the effect that knowledge on agriculture may have on the decision to use a type of robot, it would be interesting to analyze the opinions of professionals who are involved at various levels of the value chains (e.g., entrepreneurs, managers, and workers).

A similar result for selective activities, both high-risk and low-risk, was found for the humanoid robot compared to other types of robots. In this case, it can be hypothesized that the level of anthropomorphization had an effect on people's opinion about the use of robots in agricultural activities.

Taken together, the results on the human and the humanoid robot may suggest that anthropomorphization has little effect on the comparison between humans and robots since they are placed on two different ontological levels; however, it has an effect within the same ontological level, as for robots. In particular, it is possible to hypothesize a driving effect of robotic anthropomorphization on people's opinions concerning their use in different agricultural activities. This reflection is further supported by the differences found between the Autonomous Vehicle and the Quadrupedal Robot with respect to their use in Non-Selective/High-Risk activities, with autonomous vehicles producing the higher score. However, it is important to underline that observing the actual functioning of the robots in their operational environment (i.e., by sharing short videos) could lead to an increased understanding of the real technological potential of each robotic platform and, as a consequence, affect people's perceptions on their use in agriculture. This last issue should merit further investigation in future studies. Furthermore, despite the differences found in the use of robots in agricultural activities between hu-

man and robots and between different types of robots, it is important to note that this preliminary study depicts an interesting context where people (potentially representing the consumer category) would accept the use of robotic technologies in agriculture. This finding should be further investigated as it could demonstrate a general tendency in our society to accept robotics in the food value chain, demonstrating a shift from the widely held paradigm of agriculture as manual labor-dependent, to an integrated human-robot model.

Regarding people's opinions on the value of robots compared to humans in different agricultural processes, the results showed that the humanoid robot is considered better than the other robotic agents for selective tasks by maintaining high quality standards and safety, while the autonomous vehicle is considered better than the Quadruped Robot for the Productivity domain and the humanoid robot for the work domain. Once again, the results show how the robot's anthropomorphization influences people's opinions. Regarding the usefulness of robots in agricultural activities, the humanoid robot is preferred to the other robots. It can be argued that the greater attribution of mental qualities to these robots positively affects people's opinions of their perceived safety and quality in selective activities. This result is partly explained by agricultural knowledge. Specifically, people with higher levels of agricultural knowledge are also those who evaluate humanoid robots better for the safety domain compared to the other robots. This finding shows that agricultural knowledge also influences the value associated with the robot. For the productivity domain, people attribute a higher value to the autonomous vehicle than to the quadruped robot. This result might reflect a classical view of the use of technologies in agriculture, i.e., the tractor as a means to improve farming productivity. Finally, the autonomous vehicle represents more of an opportunity in agricultural work compared to the humanoid robot; although rated less useful than the human in various activities (see reflections above), the humanoid robot could nevertheless present a greater threat than the autonomous vehicle to human labor in agriculture. The idea that may underpin this finding is plausibly related to the greater tendency to assimilate the humanoid robot with the human, as evidenced by the greater propensity to ascribe mental states to humanoid robots compared to other robots. Consequently, the humanoid robot could be considered as a greater competitor of humans.

Conclusions

To summarize, the present study investigated people's perceptions of the use of robots in agriculture and how they evaluate the impact that

robots might have in different domains of agriculture. The study analyzed the human perceptions of a 49-people panel comprising mainly students enrolled in the Faculties of Education and Psychology with a moderate knowledge of agriculture, only 13.2% of those interviewed claiming a more solid background. In general, the results show that people prefer the use of humans over robots to perform selective agricultural activities; probably, this finding can be explained by the greater reliability associated with humans in making decisions and performing selective tasks. On the other hand, when comparing the different robots, the humanoid robot 'takes the place' of the human, i.e., people consider it as the robot to be used in selective activities requiring more complex decisions. This last result is also confirmed when analyzing the perception on the effect that the introduction of humanoid robots has in different domains of agriculture: there is a preference for the use of humanoid robots for selective tasks in both the safety and quality domains. Finally, autonomous vehicles are preferred in terms of increased work productivity and reduced risk.

Generally, these findings allow us to hypothesize that people place humans and robots on two different ontological levels with respect to their mental characteristics, as evidenced by the preference for using robots in tasks that do not require complex decisions. This consideration is further supported by the limited differences found between humans and robots in both high- and low-risk non-selective tasks.

Overall, the results show the importance of investigating the emerging topic of robotics in agriculture also on a psychological perspective. Obviously, the study is not without limitations (e.g., sample size and a prevalence of students from Education); however, the study provided promising new lines of research in the agricultural and psychology fields.

References

Baxter, P., Ashurst, E., Read, R., Kennedy, J., & Belpaeme, T. (2017). Robot education peers in a situated primary school study: Personalisation promotes child learning. *PloS One*, *12*(5), e0178126.

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science robotics*, *3*(21).

Botterill, T., Paulin, S., Green, R., Williams, S., Lin, J., Saxton, V. et al. (2017). A robot system for pruning grape vines. *Journal of Field Robotics*, *34*(6), 1100-1122.

Casas, J., Senft, E., Gutierrez, L. F., Rincon-Rocancio, M., Munera, M., Belpaeme, T., & Cifuentes, C. A. (2020). Social assistive robots: Assessing the impact of a training assistant robot in cardiac rehabilitation. *International Journal of Social Robotics*, 1-15.

Cavallo, F., Esposito, R., Limosani, R., Manzi, A., Bevilacqua, R., Felici, E., Di Nuovo, A., Cangelosi, A., Lattanzio, F., & Dario, P. (2018). Robotic services acceptance in smart environments with older adults: User satisfaction and acceptability study. *Journal of medical Internet research*, *20*(9), e264.

Colgate, J., Wannasuphoprasit, W., & Peshkin, M.A. (1996). Cobots: Robots for collaboration with human operators. In *Proceedings of the international mechanical engineering congress and exhibition*, 58, pp. 433-439.

Di Dio, C., Manzi, F., Itakura, S., Kanda, T., Hishiguro, H., Massaro, D., et al. (2019). It does not matter who you are: Fairness in pre-schoolers interacting with human and robotic partners. *International Journal of Social Robotics*, *20*(5), 1-15. doi: 10.1007/s12369-019-00528-9.

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., et al. (2020a). Come i bambini pensano alla mente del robot. Il ruolo dell'attaccamento e della Teoria della Mente nell'attribuzione di stati mentali ad un agente robotico. *Sistemi Intelligenti, 32*, 41-56. doi: 10.1422/96279.

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., et al. (2020b). Shall I trust you? From child human-robot interaction to trusting relationships. *Frontiers in Psychololog*y, 11:469. doi: 10.3389/fpsyg.2020.00469.

Gatti, M., Squeri, C., Garavani, A., Vercesi, A., Dosso, P., Diti, I., & Poni, S. (2018). Effects of variable rate nitrogen application on cv. barbera performance: Vegetative growth and leaf nutritional status. *American Journal of Enology and Viticulture*, 69(3), 196-209.

Kootstra, G., Wang, X., Blok, P. M., Hemming, J., & Van Henten, E. (2021). Selective harvesting robotics: Current research, trends, and future directions. *Current Robotics Reports*, 1-10.

Manzi, F., Massaro, D., Kanda, T., Kanako, T., Itakura, S., & Marchetti, A. (2017). Teoria della Mente, bambini e robot: L'attribuzione di stati mentali. *Proceedings of XXX Congresso Nazionale AIP, Sezione di Psicologia dello Sviluppo e dell'Educazione*, Messina (September 14-16), 65.

Manzi, F., Peretti, G., Di Dio, C., Cangelosi, A., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., & Marchetti, A. (2020). A robot is not worth another: Exploring children's mental state attribution to different humanoid robots. *Frontiers in Psychology*, *11*, 2011.

Marchetti, A., Di Dio, C., Manzi, F., & Massaro, D. (2020). Robotics in clinical and developmental Psychology. *Reference Module in Neuroscience and Biobehavioral Psychology, B978-0-12-818697-8.00005-4*. https://doi.org/10.1016/B978-0-12-818697-8.00005-4.

Matheson, E., Minto, R., Zampieri, E. G., Faccio, M., & Rosati, G. (2019). Human-robot collaboration in manufacturing applications: A review. *Robotics*, *8*(4), 100.

Oliveira, L. F., Moreira, A. P., & Silva, M. F. (2021). Advances in agriculture robotics: A state-of-the-art review and challenges ahead. *Robotics*, *10*(2), 52.

Poni, S., Tombesi, S., Palliotti, A., Ughini, V., & Gatti, M. (2016). Mechanical winter pruning of grapevine: Physiological bases and applications. *Scientia Horticulturae,* (204), 88-98.

Sparrow, R., & Howard, M. (2021). Robots in agriculture: Prospects, impacts, ethics, and policy. *Precision Agriculture, 22*(3), 818-833.

Vougioukas, S. G. (2019). Agricultural Robotics. *Annual Review of Control, Robotics, and Autonomous Systems*, 2, 365-392.

Williams, H., Nejati, M., Hussein, S., Penhall, N., Lim, J. Y., Jones, M. H. et al. (2020). Autonomous pollination of individual kiwifruit flowers: Toward a robotic kiwifruit pollinator. *Journal of Field Robotics, 37*(2), 246-262.

SECTION 4

Towards a Humane Technology
Challenges and Perspectives

1. The Neuroscience of Smart Working and Distance Learning

G. Riva, B.K. Wiederhold, F. Mantovani

ABSTRACT

The persistence of the coronavirus-caused respiratory disease (COVID-19) and the related restrictions to mobility and social interactions are forcing a significant portion of students and workers to reorganize their daily activities to accommodate the needs of distance learning and agile work (smart working). What is the impact of these changes on the bosses/teachers' and workers/students' experience?

This article uses recent neuroscience research findings to explore how distance learning and smart working impact the following three pillars that reflect the organization of our brain and are at the core of school and office experiences: a) the learning/work happens in a dedicated physical place; b) the learning/work is carried out under the supervision of a boss/professor; and c) the learning/work is distributed between team members/classmates. For each pillar, we discuss its link with the specific cognitive processes involved and the impact that technology has on their functioning. In particular, the use of videoconferencing affects the functioning of Global Positioning System neurons (neurons that code our navigation behavior), mirror neurons, self-attention networks, spindle cells, and interbrain neural oscillations. These effects have a significant impact on many identities and cognitive processes, including social and professional identity, leadership, intuition, mentoring, and creativity. In conclusion, just moving typical office and learning processes inside a videoconferencing platform, as happened in many contexts during the COVID-19 pandemic, can in the long term erode corporate cultures and school communities. In this view, an effective use of technology requires us to reimagine how work and teaching are done virtually, in creative and bold new ways.

This chapter was originally published as Riva, G., Wiederhold, B.K., & Mantovani, F. (2021). Surviving COVID-19: The neuroscience of smart working and distance learning. *Cyberpsychology, Behavior, and Social Networking, 24*(2), 79-85. No competing financial interests exist. The preparation of this article was supported by the UCSC D3.2 2020 project "Behavioural change: prospettive per la stabilizzazione di comportamenti virtuosi verso la sostenibilità".

Introduction

The persistence of the coronavirus-caused respiratory disease (COVID-19) and the related restrictions to mobility and social interactions are forcing a significant portion of students and workers to reorganize their daily activities to accommodate the needs of distance learning and agile work (smart working). Common features of these activities are as follows: a) the teacher/learner and boss/worker separation by space or time, or both; b) the learner/learner and worker/worker separation by space or time, or both; and c) and the use of media and technology to enable communication and exchange during the learning process and the working hours despite these separations. In particular, homes are becoming our schools and our offices: every day we check up on each other with online meetings, calls, and e-mails.

What is the impact of these changes on our personal experience? Individuals experiencing distance learning and smart working are now beginning to experience a new phenomenon: tiredness, anxiety, or worry resulting from the overuse of virtual videoconferencing platforms – so-called Zoom fatigue (Wiederhold, 2020a). With this term, psychologists describe a feeling of fatigue and discomfort linked to the numerous videoconferencing sessions that represent the main communicative tool used during distance learning and smart working.

Why are we experiencing this Zoom fatigue, and what is its impact on the processes of learning and working? The most common explanations are the following:

– Technology often does not work optimally. Who has not experienced connection problems or microphones and cameras that do not work? And the situation worsens, increasing the level of stress, when there are several people in the same house simultaneously online doing smart working or distance learning at the same time.

– Videoconferencing reduces nonverbal cues. In this way it requires a significant increase in cognitive resources to perceive and understand the meaning of others' communicative acts, as well as paving the way to more frequent misunderstandings (Morris, 2020). In addition, although not perceived consciously, the slight delays in videoconferencing force our brains to work harder to overcome and restore synchrony to our communications.

However, as we try to describe in this article, the scenario is more complex than it may appear at first glance. In particular, we use the recent reflections and neuroscience research findings to explore how distance learning and smart working impact the following three pillars that reflect the organization of our brain and are at the core of school and

office experiences (Bhattacharya, 2017; Gallese, 2006; Goleman & Boyatzis, 2008; Moser et al., 2015):

– Sense of Place (placeness): the learning/work happens in a dedicated physical place.

– Leadership, Empathy, Intuition, Mentoring/Scaffolding: the learning/work is carried out under the supervision of a boss/professor.

– Group Identification, Collective Performance, and Creativity: the learning/work is distributed between team members/classmates.

For each pillar, we discuss its link with the specific cognitive processes involved and the impact that technology has on their functioning (Table 1).

Sense of Place (Placeness): The Learning/Work Happens in a Dedicated Physical Space

The first thing that comes to our mind when we use the words 'office' and 'classroom' is a place where work/learning is carried out. In both cases the words describe a specific location of management and organization.

Why do we need to work and/or study in a specific place? Apart from organizational issues, the answer lies in the organization of our brain. As recently demonstrated by May-Britt Moser and Edvard I. Moser who were awarded the 2014 Nobel Prize, our brain contains different neurons – the 'place cells', 'border cells', 'grid cells', and 'head direction cells', found within the hippocampal-entorhinal circuit – which are activated when we occupy a certain position in the environment and when we identify a border within it [Global Positioning System (GPS) neurons], a specific group of neuron that, like the GPS providing geolocation to our cars, code our navigation behavior helping us to reach a destination (Moser et al., 2015).

These cells help to identify the individual's spatial coordinates and are also believed to help organize memories about specific locations (Moser et al., 2015). In particular, recent research has found that single neurons in the human entorhinal cortex change their spatial tuning to target relevant memories for retrieval (Qasim et al., 2019), suggesting that our brain collectively encodes both location information and episodic memories (Ekstrom & Ranganath, 2018) and that this process is crucial for the formation and consolidation of autobiographical memories (Boccia et al., 2019).

Autobiographical memory, the ability to recollect and re-experience personal life events occurring at a particular time and place, is critical for the development of our unique personal identity (Boccia et al.,

Table 1 - *The impact of smart working and distance learning on different brain and cognitive processes*

Pillars	*Brain processes involved*	*Cognitive processes involved*	*Impact of technology*
The learning/work happens in a dedicated physical place	GPS neurons	Autobiographical memory	Placelessness and nowhereness Reduced social and professional identity Stress (Zoom fatigue) and burnout
The learning/work is carried out under the supervision of a boss/professor	Mirror neurons Fight-or-flight response Self-attention network Spindle cells	Intentional attunement Empathy Intuition	Reduced leadership Less intuitive decisions and more cognitive load (Zoom fatigue) Difficulty in mentoring/ scaffolding
The learning/work is distributed between team members/classmates	Interbrain neural oscillations	Team engagement Social dynamics Joint attention	Reduced creativity and innovation More complex team dynamics

GPS, Global Positioning System.

2019). As noted by Lengen et al. (2019) "The ability to remember, recognize and reconstruct places is a key component of episodic autobiographical memory. In this respect, place forms an essential basis for the unfolding of experiences in memory and imagination" (p. 21).

In other words, the experience and development of identity and self are both collectively and individually anchored in the relationship to places (placeness). In fact, we define who we are through the memory of people and events that occurred within the different places we frequent: we are workers because we go every day to our office in a specific building, we are students because we go to the same classroom in the school.

However, what happens to our GPS neurons when rather than meeting in a physical locale we instead meet in videoconferencing? As demonstrated recently by Li et al. (2020), when we experience multicompartment environments (i.e., we are in a room, but simultaneously we are experiencing a digital space on the screen of the computer), our place is the space in which we can move; not the one we are viewing. In simpler words, for our brain, Zoom and Meet are not places, so they do not activate the binding of the experiences we have through them with our autobiographical memory.

Experientially the final outcome is simple: a feeling of 'placelessness'. The concept of placelessness was initially introduced by the geographer Relph to describe the loss of uniqueness of place in the cultural landscape so that one place looks like the next (Relph, 1976). In distance learning and smart working, the experience of placelessness is related to the loss of uniqueness of videoconferencing meetings inside our daily activities. All the meetings look the same and at the end of the day we feel empty and out of focus.

Moreover, the shift from physical locations (offices/classrooms) to digital locations (Zoom, Teams, Meet, etc.) places more emphasis on the quantitative rather than the qualitative aspects of the experience (Arefi, 1999). This can produce an increasing number of new meetings that do not provide a significant improvement of knowledge or work and generate a feeling of 'nowhereness': I'm always online but I'm not going anywhere.

Finally, the experience of placelessness weakens the sense of professional identity: we do not just act as a student or a worker, we are students or workers. Psychology defines professional identity as one's professional self-concept based on attributes, beliefs, values, motives, and experiences (Ibarra & Deshpande, 2007). And professional identity develops through relationships not only with people but also with places that represent settings for our professional activity (Ardoin, 2006). However, the impossibility of linking our personal activity to a place, makes us less students or workers, not only for us but also for the peo-

ple surrounding us. In fact, in our home we are not only workers but also parents, sons, nephews and so on. In practice, being outside a dedicated place – our office/classroom – does not guarantee a defined social identity that prevents family and friends from voluntarily or involuntarily interrupting our activities. Plus, it pushes us to do more things at once (multitasking) – that is, to attend the meeting while looking at the child's homework – inevitably increasing fatigue and stress.

This is why various studies have found a direct link between reduced professional identity and level of burnout (Edwards & Dirette, 2010; Devery et al., 2018): the less you feel like a student or a worker, the less others perceive it, and the more likely you are to experience burnout.

Leadership, Empathy, Intuition, And Mentoring/Scaffolding: The Learning/Work is Carried out under the Supervision of a Professor/Head

The second thing that comes to our mind when we use the words 'office' and 'school' is the peculiar relationship we have with our boss/teacher. On the one hand, they organize and control our work, by defining deadlines, priorities, and goals. On the other hand, at least for teachers, they scaffold our work, providing support and guidance for the completion of a given task. In both cases, bosses and teachers are our leaders and they provide organization and support to our activities.

Even if personal variables, and in particular personality dimensions, are effective predictors of the quality of leadership, there is a long-standing acknowledgment that situation and context can, at least in part, influence leadership behavior (Waldman et al., 2011). In particular, personality and context meet in the concept of 'social intelligence', that Goleman and Boyatzis (2008) define as: "A set of interpersonal competencies built on specific neural circuits (and related endocrine systems) that inspire others to be effective" (p. 3).

In fact, a specific type of neurons – mirror neurons – play a significant role in supporting leadership and social intelligence. Mirror neurons are bimodal neurons (i.e., that fire in responses to stimuli from more than one sensory modality) that are activated both when an individual acts and when an individual observes the same action performed by another (Iacoboni, 2006; Kaplan & Iacoboni, 2006).

In other words, the discovery of mirror neurons suggests that our brain is able to mimic, through an intuitive simulation, what another individual does and experiences. Moreover, we are able to detect someone else's emotions through their actions, because our mirror neurons produce an instant sense of attunement (Gallese, 2006). As underlined

by Gallese (2006), "A direct form of experiential understanding of others, intentional attunement, is achieved by modeling their behavior as intentional experiences on the basis of the activation of shared neural systems underpinning what the others do and feel and what we do and feel" (p. 15).

This process is critical for both bosses and teachers, because their emotions and actions prompt workers and students to mirror those feelings and deeds. In particular, as underlined by Dasborough et al. (2009), workers/students' perception of a boss/teacher's behaviors and their attribution generate a process of emotional contagion that spreads to other individuals in the group.

For example, providing positive feedback that is delivered together with negative bodily signals – frowns and narrowed eyes – generates worse feelings about the evaluated performance than does negative performance feedback accompanied by positive emotional bodily signals – nods and smiles (Goleman & Boyatzis, 2008). And these feelings are then reflected in the group's affective climate and trust climate, empathy levels, and also in the quality of leader/member and team/member relationships.

In a face-to-face setting, we perceive immediately the difference between a boss or a teacher who is self-controlled and humorless and another one who laughs and sets an easygoing tone. As demonstrated by research (Goleman & Boyatzis, 2008), top-performing leaders elicited laughter from their subordinates three times as often, on average, as did less-performing leaders. However, what happens to the same top-performing leaders when they move to smart working or distance learning?

First, in videoconferencing, only faces are visible. This implies the loss of all the cues and communicative tools coming from nonverbal bodily signals: posture and body movements, haptic communication and proxemics. This loss is further amplified by the fight-or-flight response. When we have prolonged eye contact with a large audience, our bodies get flooded with cortisol, the stress hormone. And we automatically think there is danger, even though consciously we know there is no danger.

Moreover, during a videoconference we usually see our face. On the one hand, implicit face control draws our attention away from the discussion by activating the self-attention network (Chakraborty & Chakrabarti, 2018). On the other hand, seeing our emotions amplifies their intensity, making it more difficult to control them effectively in social situations (Chakraborty & Chakrabarti, 2018).

Furthermore, many participants in videoconference meetings switch off their cameras to reduce bandwidth and/or to avoid problems with their appearance/context. Obviously, the lack of facial expressions

makes it difficult for bosses/teachers to develop an emotional bond with their team/class.

Finally, even the different nonverbal components of the message – the tone and pitch of the voice, its speed, and volume – are often distorted by low-bandwidth or connection problems.

In summary, the use of typical videoconference tools significantly disrupts the processes of intentional attunement that is based on the intuitive perception and analysis by the mirror neurons of body movements, facial expressions, and voice intonation producing a significant increase in cognitive resources needed to effectively express and understand communication (Gallese, 2006).

However, as underlined by neuroscience, another class of neurons is impacted by this situation: *spindle cells* (Kiran & Tripathi, 2018). Spindle cells, also called 'Von Economo Neurons - VENs' are special cells that drive what is termed as 'intuition' or 'gut feeling'. These cells have a large spindle-shaped body, about 4 times that of typical brain cells, to allow an ultrarapid transmission – within 1/20 of a second – of emotions, beliefs, and judgments (Gallese, 2006; Kiran & Tripathi, 2018). These cells play a critical role in leadership because they allow a boss or a teacher to immediately understand whether a student deserves a high (or low) rank or if someone is right (or wrong) for a job. Unfortunately, this process is literally physical, and it is based on the same nonverbal signals discussed before. Without access to these signals, a leader cannot use intuition to make the many quick decisions that create the difference in teamwork. Since more cognitive resources are required to make the same decisions that before were effortless, these decisions become more prone to errors.

Obviously, this situation impacts not only bosses/teachers but also workers/students. For them it is more difficult to fully understand the implicit meanings of the messages from their leaders. And in many situations how the message is communicated is more important than the message itself. Moreover, taking intuitive decisions using the information available during the videoconference is more demanding for them, too.

In particular, the lack of intentional attunement and the difficulty of taking intuitive decisions also have a strong impact on all the mentoring/scaffolding activities. With this term, we define the support given by a boss/teacher to a worker/student when performing a task that he/she might otherwise be unable to accomplish (Van de Pol et al., 2010). These experiences are based on three different characteristics (Van de Pol et al., 2010) – namely contingent responsivity, fading, and transfer of responsibility – that are strongly dependent on attunement and intuitive decisions.

'Contingent responsivity' is the ability to perceive and understand the worker/student's cues and signals related to learning, affective, and motivational needs, and then to respond in a timely and appropriate way. 'Fading' is instead the gradual withdrawal of the scaffolding and depends upon the worker/student's level of competence and skills. It is the progressive fading that allows the transfer of responsibility, that is, the performance of the task is gradually transferred to the learner.

These characteristics underline that, in mentoring/scaffolding, both participants are active participants, and they reach a common ground through a process of intuitive and progressive attunement that is based on their communicative exchanges and common work. Unfortunately, as we have just seen, this process is more difficult in smart working and distance learning, and can generate unrealistic expectations, lack of commitment, or an overdependence on the mentor/mentee that provides significant challenges for the efficacy of the process.

Group Identification, Collective Performance, and Creativity: The Learning/Work Is Distributed Between Classmates/Team Members

A final thing that comes to our mind when we use the words 'office' and 'school' is the experience of the group – the colleagues and the classmates – we come into contact with. In fact, work/study is always interconnected. So, an effective interpersonal work/school relationship – helping others and/or receiving support from them – is a critical factor for the success and satisfaction with job/study.

Together with mirror neurons and spindle cells, as we have seen before play a critical role in supporting social intelligence and intuition, there is another brain activity that is involved in these processes (Gallese, 2006; Kiran & Tripathi, 2018): *neural oscillations*. Neural oscillations, or brain waves, are rhythmic patterns of neural activity that enable the coordinated activity of the brain (Cebolla & Cheron, 2019). In particular, neural oscillations in the high frequencies (gamma and beta) and in the low ones (alpha and theta) allow accurate temporal synchrony between distributed neuronal responses.

The dynamics of neural oscillations is not influenced only by the individual's interaction with the physical environment. A growing body of studies using a new brain imaging technique – hyperscanning (Balconi & Vanutelli, 2017), which allows measurement of the activity of multiple brains simultaneously – has revealed that neuronal oscillations are also influenced by social dynamics (Balconi & Vanutelli, 2018; Kinreich et al., 2017). In particular, natural social interactions produce a brain-to-brain synchrony in neural oscillations during natural social interac-

tions that is linked with behavioral synchrony. In other words, synchronized neuronal oscillations coordinate people physically by regulating how and when their bodies move together (Lindenberger et al., 2009). As explained by Goleman and Boyatzis (Gallese, 2006), "You can see oscillators in action when you watch people about to kiss; their movements look like a dance, one body responding to the other seamlessly. The same dynamic occurs when two cellists play together. Not only do they hit their notes in unison, but thanks to oscillators, the two musicians' right brain hemispheres are more closely coordinated than are the left and right sides of their individual brains" (p. 4).

Starting from these premises, social neuroscience recently provided a clear answer to a critical question for our goals: what happens in our neural oscillations when we work together with our team or our classmates?

In a first study, Reinero et al. (2021) explored the effects of collective behavior within small teams (four persons) using electroencephalography (EEG) hyperscanning to simultaneously record each person's brain activity. Specifically, they divided 174 participants in groups of 4 asking them to complete a series of problem-solving tasks either independently or as a team. On the one hand, teammates identified more strongly with one another, and outperformed the average individual on most problem-solving tasks. On the other hand, interbrain synchrony was the best predictor of collective performance.

In a second study, Dikker et al. (2017) used EEG hyperscanning to evaluate the potential link between classroom social dynamics – a class of 12 students in their senior year at a high school in New York City – and interbrain synchrony over a long period. Again, brain-to-brain group synchrony predicted classroom engagement and social dynamics. And in both studies, interbrain synchrony predicted collective performance better than self-report measures of group identification and behavioral indices of cooperation and emotion perception.

However, what are the factors that allow brain-to-brain group synchrony? Both studies suggest that the principal factor is 'joint attention' – the experience of two or more individuals who know that they are attending to something in common (Bhattacharya, 2017) – by tuning neural oscillations to the temporal structure of the common context. However, joint attention requires eye contact and the exchange of glances (mutual gaze; Bhattacharya, 2017): "Prior eye contact potentially creates a context for joint attention, which subsequently induces higher interbrain synchrony" (p. R347).

However, what happens to joint attention and interbrain synchrony when we move to smart working or distance learning?

Even if faces are usually visible in videoconferencing, eye contact and

exchange of glances are not possible: you cannot look simultaneously into the camera and at the faces on the screen. This implies that the development of joint attention is difficult and requires more complex and less intuitive strategies. Moreover, it can impact creativity and innovation, which are strongly dependent on collective performance and on interactions within a social network. As underlined by Glăveanu and Lubart (2014), even creative individuals use social interaction for elaborating, championing, and implementing their ideas. For example, social interaction is used by them for identifying constraints needed to guide their creativity by focusing it on what interests others; for energizing themselves and providing them with a sense of purpose; and for seeing how others look at things; providing them with a different perspective. However, the difficulty in creating interbrain synchrony can affect these processes and reduce engagement, brainstorming, and creativity.

Conclusions

This article started from a simple question: how does distance learning and smart working affect bosses/teachers' and workers/students' experiences?

We used recent neuroscience research to explore how distance learning and smart working impact the following three pillars that reflect the organization of our brain and are at the core of school and office experiences: a) the learning/work happens in a dedicated physical place; b) the learning/work is carried out under the supervision of a boss/professor; c) the learning/work is distributed between team members/classmates.

For each pillar, we discussed its link with the specific cognitive processes involved and the impact that technology has on their functioning. Specifically, the use of videoconferencing affects the functioning of GPS neurons, mirror neurons, self-attention networks, spindle cells, and interbrain neural oscillations with a significant impact on many identities and cognitive processes.

First, the use of technology can generate a sense of placelessness that has a direct impact on our episodic memory, our personal and professional identity, and increases the risk of burnout. More, the lack of intentional attunement and the difficulty in taking intuitive decisions have also a strong impact on leadership and on all the mentoring/scaffolding activities. Finally, the impossibility of using eye contact and the exchange of glances, the main tools used to generate joint attention, reduces group engagement, collective performance, and creativity.

These results suggest that just moving office and learning process-

es inside a videoconferencing platform, as happened in many contexts during the COVID-19 pandemic, are not an effective solution and in the long term can erode corporate cultures and school communities. In this view, an effective use of technology requires that we reimagine how work and teaching are done virtually, in creative and bold new ways.

A possible strategy to drive these efforts is to use technology to generate and support a community of practice (CoP) able to develop Networked Flow, an optimal collective experience (Gaggioli et al., 2013). According to Wenger et al. (2002), a 'community of practice' (CoP) is: "A group of people who share a concern, a set of problems, or a passion about a topic, and who deepen their knowledge and expertise in this area by interacting on an ongoing basis".

For example, the groups of adolescents meeting daily in social video games, such as Fortnite or Among Us, represent a clear example that it is possible to develop successful communities even without a physical place in which to gather together. According to Gaggioli et al. (2017), an optimal group experience is achieved by building a 'collaborative zone of proximal development', in which the actions of the individuals and those of the collective are in balance and a sense of social presence is established.

CoP is characterized by three characteristics that set them apart from a classical team and are able to overcome many of the limitations we discussed before (Paasivaara & Lassenius, 2014): a community, practice, and a domain. To overcome the lack of a common place, CoP develops a sense of community through an active engagement in joint practical activities that allow relationships with one another to develop, and information to be shared. The practical aspect means that CoPs use technological and nontechnological tools to build an organizational memory and a shared set of resources allowing them to solve problems in their domain of interest. Finally, each CoP is built around a specific area of interest in which the workers/students collaborate to share and create knowledge.

Moreover, to overcome the lack of joint attention and nonverbal cues that limit the typical videoconferencing tools, members of online CoPs use specific strategies such as action coordination (i.e., the coordination of hands and eyes in goal-directed action; Yu & Smith, 2013) and vocal exchanges (i.e., the verbal confirmation and reconfirmation that contact had been established; Sävenstedt et al., 2005). These activities generate a 'flow contagion' among CoP members, characterized by mutual comprehension, high intrinsic motivation and shared positive emotions (Gaggioli et al., 2017). This can also explain why social video games are more effective for adolescents than the typical distance learning tools in producing team engagement and group synchrony.

Finally, better technology can help us too. For example, recently Microsoft developed a 'Together Mode' for its videoconferencing tools (Basu, 2020). This mode uses artificial intelligence to take a cutout of the different live video images of the participants and place it into a fixed position within a setting (i.e., a classroom or a virtual auditorium) to increase the feeling of sharing a common space.

Another possibility is to exploit the simulative power of virtual reality (VR; Riva et al., 2019). As demonstrated recently, VR is able to activate GPS neurons (Harvey et al., 2009) and to generate empathy (Wiederhold, 2020a) allowing the development of more authentic relationships (Facebook, 2017). In line with these results, different companies are developing and/or have released recently different VR social platforms (Wiederhold, 2020b) that will be able to better support distance learning and smart working: Facebook Horizon, VIVE Sync, AltspaceVR, Spatial, and VRChat.

References

Ardoin, N. M. (2006). Toward an interdisciplinary understanding of place: Lessons for environmental education. *Canadian Journal of Environmental Education (CJEE), 11*(1), 112-126.

Arefi, M. (1999). Non-place and placelessness as narratives of loss: Rethinking the notion of place. *Journal of Urban Design, 4*(2), 179-193.

Balconi, M., & Vanutelli, M. E. (2017). Brains in competition: Improved cognitive performance and inter-brain coupling by hyperscanning paradigm with functional near-infrared spectroscopy. *Frontiers in behavioral neuroscience, 11*, 163.

Balconi, M., & Vanutelli, M. E. (2018). EEG hyperscanning and behavioral synchronization during a joint actions. *Neuropsychological Trends, 24*, 23-47.

Basu, T. (2020). Microsoft's solution to Zoom fatigue is to trick your brain. *MIT Technology Review.* https://www.technologyreview.com/2020/07/09/1004948/microsoft-together-mode-solution-to-zoom-fatigue.

Bhattacharya, J. (2017). Cognitive neuroscience: Synchronizing brains in the classroom. *Current Biology, 27*(9), R346-R348.

Boccia, M., Teghil, A., & Guariglia, C. (2019). Looking into recent and remote past: Meta-analytic evidence for cortical re-organization of episodic autobiographical memories. *Neuroscience & Biobehavioral Reviews, 107*, 84-95.

Cebolla, A. M., & Cheron, G. (2019). Understanding neural oscillations in the human brain: From movement to consciousness and vice versa. *Frontiers in psychology, 10*, 1930.

Chakraborty, A., & Chakrabarti, B. (2018). Looking at my own face: Visual processing strategies in self-other face recognition. *Frontiers in psychology, 9*, 121.

Dasborough, M. T., Ashkanasy, N. M., Tee, E. Y., & Herman, H. M. (2009). What goes around comes around: How meso-level negative emotional contagion can ultimately determine organizational attitudes toward leaders. *The Leadership Quarterly, 20*(4), 571-585.

Devery, H., Scanlan, J. N., & Ross, J. (2018). Factors associated with professional identity, job satisfaction and burnout for occupational therapists working in eating disorders: A mixed methods study. *Australian Occupational Therapy Journal, 65*(6), 523-532.

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., ... & Poeppel, D. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Current Biology, 27*(9), 1375-1380.

Edwards, H., & Dirette, D. (2010). The relationship between professional identity and burnout among occupational therapists. *Occupational therapy in health care, 24*(2), 119-129.

Ekstrom, A. D., & Ranganath, C. (2018). Space, time, and episodic memory: The hippocampus is all over the cognitive map. *Hippocampus, 28*(9), 680-687.

Facebook (2017). How virtual reality facilitates social connection. 2021, 30 January. https://www.facebook.com/business/news/insights/how-virtual-reality-facilitates-social-connection.

Gaggioli, A., Chirico, A., Mazzoni, E., Milani, L., & Riva, G. (2017). Networked flow in musical bands. *Psychology of Music, 45*(2), 283-297.

Gaggioli, A., Milani, L., Mazzoni, E., & Riva, G. (2013). *Networked Flow: Towards an Understanding of Creative Networks*. Springer.

Gallese, V. (2006). Intentional attunement: A neurophysiological perspective on social cognition and its disruption in autism. *Brain research, 1079*(1), 15-24.

Glăveanu, V. P., & Lubart, T. (2014). Decentring the creative self: How others make creativity possible in creative professional fields. *Creativity and Innovation Management, 23*(1), 29-43.

Goleman, D., & Boyatzis, R. (2008). Social intelligence and the biology of leadership. *Harvard business review, 86*(9), 74-81.

Harvey, C. D., Collman, F., Dombeck, D. A., & Tank, D. W. (2009). Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature, 461*(7266), 941-946.

Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual review of psychology, 60*, 653-670.

Ibarra, H., & Deshpande, P. H. (2007). Networks and identities: Reciprocal influences on career processes and outcomes. In H. Guntz, & M. Peiperl (Eds.), *Handbook of Career Studies* (pp. 268-282.). Sage.

Kaplan, J. T., & Iacoboni, M. (2006). Getting a grip on other minds: Mirror neurons, intention understanding, and cognitive empathy. *Social neuroscience, 1*(3-4), 175-183.

Kinreich, S., Djalovski, A., Kraus, L., Louzoun, Y., & Feldman, R. (2017). Brain-to-brain synchrony during naturalistic social interactions. *Scientific reports, 7*(1), 1-12.

Kiran, C. S., & Tripathi, P. (2018). Leadership development through change management: A neuroscience perspective. *NHRD Network Journal, 11*(4), 42-48.

Lengen, C., Timm, C., & Kistemann, T. (2019). Place identity, autobiographical memory and life path trajectories: The development of a place-time-identity model. *Social Science & Medicine, 227*, 21-37.

Li, T., Arleo, A., & Sheynikhovich, D. (2020). Modeling place cells and grid cells in multi-compartment environments: Entorhinal-hippocampal loop as a multisensory integration circuit. *Neural Networks, 121*, 37-51.

Lindenberger, U., Li, S. C., Gruber, W., & Müller, V. (2009). Brains swinging in concert: Cortical phase synchronization while playing guitar. *BMC Neuroscience, 10*(1), 1-12.

Morris, B. (2020). Why does Zoom exhaust you? Science has an answer. *Wall Street Journal.* https://www.wsj.com/articles/why-does-zoom-exhaust-you-science-has-an-answer-11590600269.

Moser, M. B., Rowland, D. C., & Moser, E. I. (2015). Place cells, grid cells, and memory. *Cold Spring Harbor perspectives in biology, 7*(2), a021808.

Paasivaara, M., & Lassenius, C. (2014). Communities of practice in a large distributed agile software development organization - Case Ericsson. *Information and Software Technology, 56*(12), 1556-1577.

Qasim, S. E., Miller, J., Inman, C. S., Gross, R. E., Willie, J. T., Lega, B., ... & Jacobs, J. (2019). Memory retrieval modulates spatial tuning of single neurons in the human entorhinal cortex. *Nature neuroscience, 22*(12), 2078-2086.

Reinero, D. A., Dikker, S., & Van Bavel, J. J. (2021). Inter-brain synchrony in teams predicts collective performance. *Social Cognitive and Affective Neuroscience, 16*(1-2), 43-57.

Relph, E. (1976). *Place and Placelessness (Vol. 67)*. Pion.

Riva, G., Wiederhold, B. K., & Mantovani, F. (2019). Neuroscience of virtual reality: From virtual exposure to embodied medicine. *Cyberpsychology, Behavior, and Social Networking, 22*(1), 82-96.

Sävenstedt, S., Zingmark, K., Hydén, L. C., & Brulin, C. (2005). Establishing joint attention in remote talks with the elderly about health: A study of nurses' conversation with elderly persons in teleconsultations. *Scandinavian Journal of Caring Sciences, 19*(4), 317-324.

Van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher-student interaction: A decade of research. *Educational psychology review, 22*(3), 271-296.

Waldman, D. A., Balthazard, P. A., & Peterson, S. J. (2011). Social cognitive neuroscience and leadership. *The Leadership Quarterly, 22*(6), 1092-1106.

Wenger, E., McDermott, R. A., & Snyder, W. (2002). *Cultivating Communities of Practice: A guide to Managing Knowledge*. Harvard Business Press.

Wiederhold, B. K. (2020a). Connecting through technology during the coronavirus disease 2019 pandemic: Avoiding 'Zoom Fatigue'. *Cyberpsychology, Behavior, and Social Networking, 23*, 437-438.

Wiederhold, B. K. (2020b). Embodiment empowers empathy in virtual reality. *Cyberpsychology, Behavior, and Social Networking, 23*, 725-726.

Wiederhold, B. K. (2020c). Beyond zoom: The new reality. *Cyberpsychology, Behavior and Social Networking, 23*, 1-2.

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PloS One, 8*(11), e79659.

2. Critical Thinking in the Data Age

New Challenges, New Literacies

P.C. Rivoltella

ABSTRACT

The chapter starts from a presentation of the current situation of communication, marked by the so-called 'fourth wave' of media development. It is characterized by a pervasive presence of media and the role of platforms. This means that consumer data are used by platforms owners in a situation of asymmetry between them. This situation affects our way of thinking about Media Education. In this regard, it is necessary to have in mind a bigger picture than that which normally leads us to believe that Media Education means to control the use of devices. Today, Media Education is above all Data Literacy. This forces to find new tools for the critical analysis of texts and for the exercise of active citizenship. The second part of the chapter tries to show the way also through some case studies.

Data and the Platform Society

Today we are experiencing what we can call the 'fourth wave' of media development, starting from the rise of the industrial press (Colombo, 2020).

The first wave brought us industrialization, the first steps towards mass society and the birth of public space as we knew it in the '90s, until digital and social media began to transform it. This was the age of the first printed newspapers and *feuilleton*s.

The second wave brought us audiovisual media. After the invention of the telegraph in the XIX century paved the way, audiovisual media took shape in the great season of cinema, radio and television. This was a time when content ruled, and the power of the media was affirmed, both in its adoption by totalitarian states between the wars[1] and then by postwar democracies, as numerous examples could demonstrate.

[1] For Mussolini, radio was a "very fascist instrument"; Goebbels broadcast the first experimental television programs of the Reich; Lenin understood the extraordinary power of cinema as a means of educating the masses.

The third wave brings us computerization. The internet is king. It represents the birth of a new, more horizontal way of communicating, built on disintermediation (Missika, 1983): a process that redefines the media economy but also the rules for accessing public space as well as the concept of public space itself.

We are thus in the fourth wave, marked, as we anticipated, by mediatization and the central role of data and platforms.

The idea of mediatization is different from that of simple mediation (Rath, 2017). The latter defines the classic function of the media, that is 'to mediate'. Media act as transmission devices and mediators of the main human experiences: knowledge building and sharing, memory storage and retrieval and relationship management (Thompson, 1995). A mediatized society, on the other hand, is a society in which the diffusion of the media is so widespread that it permeates any individual and social practice: to use an image that is now perhaps overused, all this can be expressed by saying that in a mediatized society media are 'onlife' (Floridi, 2014).

This media society is also a platform society (van Dijck, Poell, & de Waal, 2018), that is a society made up of programmable architectures (i.e., Facebook, Amazon, Netflix and Airbnb) whose purpose is to organize interactions between users, whilst seeking hegemony in the market: this is possible thanks to their mediation between users and the services offered to them. This goal is pursued thanks to powerful algorithms that allow platforms to use user-related data both to improve their performance in terms of services and to generate economic value.

Brunton and Nissenbaum (2015) better clarify this aspect by specifying that there are three types of data on which basis platforms work.

First, the data that users themselves provide media with. This is the case with the content that each of us uploads to our social media profiles: stories, posts, images, any kind of information. We also have the data entered in app registration forms. Finally, there is the log-in data for sites, restricted areas and Learning Management System. In all these cases, nothing is extorted by deception: it is data that we know we are leaving on the net, and without which we would not even be able to use the applications and environments to which they pertain.

A second type of data are those obtained by users for the benefit of others. This is the case with data mining, that is the extraction, analysis and classification of data based on large clusters of information. In this case, we never authorize this type of work to be carried out using data that we did not even know were part of the cluster from which the extraction is made.

Finally, we have the data obtained by users for their own benefit. This is the case for all dashboards displaying the times and methods of indi-

vidual uses of digital media such as, among others, Apple's mobile service. This third type of data seems to have greater prosocial value, since it seems that they are used in the interest of the individual user: but if a device can provide me with my biometric record on a weekly basis or a diagram representing my smartphone use, then the platform owns these data and therefore can easily make them available to third parties.

The pervasiveness of this mechanism, making it impossible for us not to be tracked, has led some to talk of a new form of capitalism organized around the value of data and the role of algorithms (Eugeni, 2021): a 'surveillance capitalism' (Zuboff, 2019) within which everyone is traceable, the result of a new revolution in the way value is realized in the economic field, where it is, indeed, equated with information.

The 'Bigger Picture': Data, Between Utopia and Critical Awareness

If we look at this situation in terms of media literacy and media literacy education, we immediately understand the need for a paradigm shift, a new perspective on the challenges that platforms pose and how these can be intercepted.

Media literacy education currently seems to be undergoing a process of normalization (Rivoltella, 2020) in at least two senses.

Media literacy education is 'normal' in the sense that Kuhn (1962) speaks of normal science in his interpretation of the historical development of scientific theories. At this stage, a new theory is imposed, recognized and accepted and, at least for some time, should not be subjected to criticism. So, the normal science phase is the quietest one, because it is based on the hegemony of a scientific theory, but it is also the least an interesting one, precisely because the work of those who try to test it is less assiduous. This is what is happening today with media literacy: people are talking about it, and it seems that social awareness of media and the need to educate people in this regard is a widespread issue in schools, families and informal education spaces.

Media literacy education is also 'normal' because it is light. It is light when it offers simple solutions, when it relies on guidelines and rules, when it passes through commercial advertising disguised as social advertising, when it relies on parental control. Common to all these proposals is the idea that media can be tamed and that this depends on the user's ability to be in control of their devices. It is precisely this idea that David Buckingham argues against in his recent Media Education Manifesto (2019), proposing a 'bigger picture' to explain what media literacy education does not explain today. The crucial question is to divert attention from the tools, their use and consumption time, to under-

standing the role of datafication and digital capitalism or GAFA (Google, Apple, Facebook, Amazon) capitalism. To believe that media education equates to controlling how much time children spend with a smartphone in their hands, or to imagine that a parental control device is enough to protect against the risk of exposure to unsuitable content, means playing the game of these large commercial players who, not surprisingly, support this type of perception and awareness: in fact, iPhone updates users on their weekly-use statistics, and Google's Family Link allows parents to monitor the duration and type of consumption by their children, to locate them and to block the use of certain apps if needed. These are corporate social responsibility operations which, while contributing to the image of an ethical and concerned company, help sustain that precise lightness mentioned above by directing the attention of educators to the tools. In this way, the underlying economic principle, datafication with all its implications, remains out of the spotlight of public opinion.

The result they intend to achieve is to opacify the processes, or to make the fact that platforms collect our data and use it invisible (or unnoticeable). Here is one of the challenges that the bigger picture presents media literacy education, that is data awareness, with today: making users aware that they are in a datafied world where the media are not (only) devices but the logic underlying them. In this type of context, a critical approach is required not to the content or functioning of the tool, but to the entire economic system (with its political motivations) of which the platforms are only the terminals.

To this argument – the need to think about a bigger picture than that provided by 'light' media education – it is possible to add further considerations: they help to better understand the importance of reflecting on data and their social role.

A first issue to consider related to what Brunton and Nissenbaum (2015) call 'information asymmetry'. This term indicates a situation in which the user, even if they know their data is used by the platforms, is deprived of the understanding of how their data are collected and used and for what purposes.

In addition to this, datafication raises two further issues (Boyd & Crawford, 2012).

The first is an epistemological one: data do not speak for themselves and always need someone to interpret them. Accordingly, it is the myth of the objectivity and definitiveness of data that must be deconstructed. This is big data hype, a perspective that leads to a sort of data fetishism, whereby data are treated "as a superior form of evidence" (Battista & Conte, 2017; p. 147) and as "an absolute, a single ahistorical artifact that speaks from a decontextualized place of authority and is alone

in providing 'real answers' to social questions" (Papas, Emmelhainz, & Seale, 2016; p. 179).

The other issue is of an ethical nature: it is not certain that, if data can provide useful information to those who use them, it is also right to use this information. Without going into depth about this very complex issue, in the context of health data, we can consider the use of data in medicine and the relationship of this use with the patient's right to informed consent and privacy.

Finally, on a psychodynamic or anthropological level, we must consider how data are able to modify the perception that each one of us has of our own self: this means that our quantified self can influence our perception of who we are and how we are perceived by others. As Carrington (2018) points out, the identity that algorithms construct can influence the way in which each person perceives themselves due to the "gap between who we think we are and who algorithms construct us to be" (2018; p. 71).

Data Literacy

From what has been said thus far, it is clear that datafication represents a problem for education. There are at least two principles that lead to the development of data literacy.

The first corresponds with a functionalist paradigm: it is the classic theme according to which in an informational society the employability of people is directly proportional to their skills, leading to perfect integration with the social and productive system and therefore to developing knowledge and skills related to the data itself. Projections on the labor market lead in this direction, as the Future Jobs survey promoted by the World Economic Forum demonstrates. It predicts that by 2025 the emerging professions will grow by 13.5%: these professions include those in cloud computing, big data and e-commerce. New emerging job profiles that will find more and more roles in companies include e-commerce and social media specialists, data analysts, AI and machine learning specialists, software and application developers. In response to this predicted trend, higher education providers have already chosen to incorporate more statistics and data analysis into their degrees (as, more than elsewhere, degree courses in the economic field clearly demonstrate). And the focus on coding and computational thinking, in both primary and secondary schools, also echoes this trend.

Behind this first principle is a precise way of thinking about media literacy in relation to data. It consists in the assumption that data literate individuals:

– have developed 'analytical thinking' as the main strategy for approaching problem solving;

– are experienced in understanding and manipulating the data produced by platforms.

The limitation of this perspective lies in the belief that the main problem is managing data, above all in a decontextualized way. On the contrary, data must always be considered as part of a broader sociomateriality, as texts that can only ever be interpreted in light of their context: essentially, we must consider data literacy within the context of new literacies (Boyd & Crawford, 2012).

The second principle belongs to a critical paradigm: it responds to that which the first one could not explain. The problem of data literacy, according to this approach, is not to encourage social adaptation and work placement, but to promote people's emancipation and their autonomy from the economic and political motivations underlying the datafication process. The switch is from data literacy to critical data literacy, of which Špiranec, Kos, and George (2019) indicate the five main dimensions:

– the ontological treatment of data (data are not transparent; they must always be placed in their context; they must not be considered in absolute terms, but are always subject to interpretation);

– criticism of the epistemological status of data (data do not represent an indisputable authority; they provide a reductive vision of reality; they can be considered a real form of ideology);

– the areas of a data literacy rationale (they have to do with algorithms, big data, data science, artificial intelligence and the neoliberal logic that governs markets and their related policies);

– the pedagogical and ethical dimensions.

Understood in the light of this critical paradigm, data literacy is thus placed in the context of the other literacies, composing a useful framework that is also applicable to the curriculum. This framework can include:

– information literacy (refers to everything concerning the retrieval, analysis, use and sharing of information);

– digital literacy (relates to the system of skills that regulate people's relationship with digital media and which finds expression at an international level in the DiGComp framework and other models developed for the same purpose);

– media literacy (underlying all the different literacies and consisting of the critical analysis of media content).

In this context, data literacy

> includes the ability to read, work with, analyze and argue using data […]. Reading data involves understanding what data is, and what aspects of the world it

represents. Working with data involves creating, acquiring, cleaning, and managing it. Analyzing data involves filtering, sorting, aggregating, comparing, and performing other such analytic operations on it. Arguing with data involves using data to support a larger narrative intended to communicate a certain message to a particular audience (D'Ignazio & Bhargava, 2015, p. 2).

Thinking Critically about Data from the Perspective of Citizenship: A Proposal for the Curriculum

One last step remains, namely to ask ourselves how data literacy as outlined above can be translated into curricular terms (in the case of formal education contexts such as school) and guidance for practice (in the case of informal educational contexts, such as adult education or risk prevention in adolescence). In other words, once the meaning of data in our social system has been clarified and the challenges they pose to education identified, how can we train people to interface correctly with them?

The general perspective in this regard is certainly that of an emancipatory paradigm that sees data literacy as a tool for promoting critical awareness and active citizenship. Many authors, including D'Ignazio and Bhargava (2015), suggest returning to Paulo Freire to establish critical data analysis and consequently develop a critical approach to data literacy. Applying the Freirean concept of conscientization, these authors identify four fundamental issues on which a critical educational intervention on data should focus:

– the lack of transparency (people are often not aware either of the fact that their data is collected, or of what is done with them);

– data is usually collected by third parties (individuals are consequently excluded from the possibility of playing an active role);

– the technological complexity of the analyzes conducted (data analysis tools and techniques are sophisticated and complex and beyond the understanding of unskilled people);

– impact control (it is impossible for people to control the consequences of the use of their data).

How can all this be translated into a framework to be adopted for curricularization and educational planning?

The literature is rich in proposals in this regard, such as the Personal Data Literacies (PDL) Framework by Pangrazio and Selwyn (2019), certainly one of the clearest and didactically adaptable of the many that it is possible to analyze. This framework is based on five areas corresponding to an equal number of actions that people are asked to perform on data, with a key question each one of the areas themselves: identification, understanding, reflexivity, use and tactics (see Table 1).

Table 1 - *The personal data literacies framework*

Area	*Key Questions*	*Actions*
Data identification	What are data?	Identification of personal data and their type (materialization).
Data understandings	What are the origins, circulations and uses of these different types of personal data?	Identifying how and where personal data are generated and processed (data trails and traces). Interpreting the information that is represented by processed data (data visualizations, charts and graphs).
Data reflexivity	What are the implications of these different types of personal data for me and for others?	Analyzing and evaluating the profiling and predictions that are made from processed personal data (i.e. sentiment analysis, natural language processing). Understanding the implications of managing, controlling and applying personal data (individual and collective critique).
Data uses	How can I manage and make use of these different types of personal data?	Applying, managing and controlling data-building technical skills and interpretive competencies (reading the terms and conditions, adjusting privacy settings, blocking technologies, developing a shared language). Applying the information represented by processed data (personal insights into digital self and performance).
Data tactics	How can I do personal data differently?	Employing tactics of resistance and obfuscation (tactics). Repurposing data for personal and social reasons (creative applications).

Each of the individual areas of the framework can be taught through different activities. Let us take some examples of these activities related to primary school, which are part of the workshops usually held in schools by CREMIT educators[2]. Specifically, the activities we briefly present are the result of the trialing of a vertical curriculum on media education in the first cycle of education[3] conducted in the Milan metropolitan area over the last three years.

Media menu

The 'Media Menu' activity consists in asking students to recognize and contextualize the data in their possession regarding their media uses (which tools, for how long, which programs, etc.) and to convert them into an infographic in order to communicate them to classmates and teachers. The two skills that this workshop aims to develop are the ability to:

– decode and communicate data, to develop in students the ability to analyze the reality that surrounds them (data identification);

– promoting the student's ability to visualize, interpret and manipulate data, representing them in different ways and increasing their own critical-reflective thinking skills (data reflexivity).

To share or not to share? This is the question!

For this activity, the classroom is divided into two areas: the first represents the public space, in which to exhibit artifacts created by the students within a frame in the external corridor; the second area, on the other hand, represents the private space, in which materials that are not intended to be shared are stored inside a suitcase (Figure 1).

The goal is:

– to encourage students' awareness of the possibility of managing their data (data uses);

– to make them responsible for their choices in terms of their potential consequences (data reflexivity).

---

[2] CREMIT (Research Center on Media Education, Innovation and Technology) is one of the research centers of the Università Cattolica del Sacro Cuore. Founded in 2006, it carries out research and teacher training in media education, education technology and innovation in school and training contexts. The examples we report are the result of the work of MELAB (Media Education LAB), which, within CREMIT, specifically deals with the development of educational tools and activities for schools. Specials thanks to Cristina Garbui and Giorgia Mauri, two of the educators involved in creating these activities.

[3] The first cycle of education in Italy spans ages 6-14 and is made up of primary school (6-11) and lower secondary school (12-14).

Figure 1 - *The private space suitcase*

The privacy thermometer

For this activity, the class puts up a colored rope, blue at one end, fading to white at the other; it traverses and connects the private dimension and the shared, public one (Figure 2). Students are asked to take pictures and hang them on the rope after reflecting and deciding whether the individual pictures should be placed on the public or private part. To each picture, students are asked to attach a post-it with three pieces of information: the target, the place within which to share the artifact and the reasons behind the decision to share or not share the image.

The goal of the whole activity is to raise children's awareness of how their personal data is used within the social universe (data understanding, data uses).

Figure 2 - *Pictures hanging on the public-private rope*

Cookieopoly

The activity offers students a reinterpretation of the well-known *Monopoly* board game (Figure 3). The rules and mechanisms of the game are reimagined through the lens of data. Thus, the land to be purchased is transformed into digital places (such as Wikipedia or Spotify), and the train stations into online newspapers; economic capital is datafied, and so banknotes represent the market research in which we participate or the personalization of our advertisement activities; unexpected events and chances relate to everyday experiences in data society (identity theft, electronic fraud, etc.).

Students immerse themselves in the game and, while playing, becomes aware of what it means to give up their data and, above all, come to understand why owning data makes it possible to become rich.

The goal of the game is to make children understand the dynamics underlying the sharing of personal data within digital platforms and to teach them to manipulate these same dynamics, albeit in a playful way.

To Conclude: Training for Participation

We now come to the end of our analysis, which saw us describe three steps.

First, it made us aware of the situation in which we find ourselves today with regard to the media industry, marked by the so-called 'fourth wave' of media development: data are at the heart of this scenario, and

Figure 3 - *Cookieopoly's dashboard*

we have had the opportunity to briefly outline their typologies and pervasiveness.

This first step then allowed us to reflect on how the data landscape described challenges education, pushing us to rethink the meaning of media literacy. We let ourselves be guided by the idea of a 'bigger picture' helping us understand the problem posed by the use of tools in a broader scenario, in which the great economic and political forces affecting digital capitalism are the real players.

The point of arrival was to identify a framework for data literacy and to try, on this basis, to think about how to teach it. The result of this work was to identify spaces within educational contexts in which we can raise people's awareness, helping them live in an increasingly critical and autonomous way.

Therefore, as we can see from this last point, it is impossible today not to consider data literacy as a fundamental element of our citizenship. Pawluczuk et al. (2020) encapsulate this by building their Data Citizenship (DC) Framework, the three dimensions of which – thought, ac-

tion and participation – allow us to understand the five areas of the PDL Framework and come to some conclusions.

Table 2 - *The Data Citizenship Framework in reference to the PDL Framework*

Area	*Actions*	*PDL Framework*
Data thinking	Understanding data collection and data economy.	Data identification, understandings, reflexivity.
Data doing	Deleting data and using data in an ethical way.	Data uses, tactics.
Data participation	Taking proactive steps to protect individual and collective privacy and wellbeing in the data society as well as helping others with their data literacy.	

As the Table 2 shows, the five areas of the PDL Framework cover only the first two areas of the DC Framework, namely those identify thinking about data and dealing with them: the dimension of participation remains external. We can refer to Pawluczuk et al. (2020, p. 15) to understand what this dimension means:

> To reflect the idea of proactive and potentially critical digital literacy, we sought to have a better understanding of how citizens participate. In our data citizenship model, we called these activities data participation. This dimension focuses on how citizens participate and especially those connections between practices which integrate online and offline activities and how they inform each other. Data participation therefore helps us to examine the collective and interconnected nature of data society. Through data participation citizens can seek opportunities to exercise their rights and to contribute to and shape their collective data experiences.

This participatory dimension insists on the intersubjective and proactive scope of data literacy. This is an important reminder because traditional media education programs often lack this dimension. So, media education is often reduced to an individual exercise. Moreover, if media education becomes just another subject on the curriculum (alongside History, Math and so on), it loses much of its power as a device for training active citizenship.

Secondly, working on the value of participation also means starting to focus on overcoming a merely deconstructive media education. Traditional programs are stuck in this setting, reducing the development of critical thinking to media-product analysis; but this work, in the case of

data, not only becomes much more complicated to carry out, but risks not being effective if it fails to inspire media-active practices and behaviors.

Understanding this also means understanding that in the data society, alongside criticism and responsibility, resistance must also form the basis of media literacy.

Critical thinking has been a core element of media literacy since the age of traditional media: it still represents its key pillar today, even in the data society. Through critical thinking, we can be aware of how the media work by exercising suspicion towards their opacity, as already suggested by the first media education researchers (Masterman, 1985).

The switch to new media and social media requires us to pair critical thinking with responsibility: viewers also become authors and, when publishing their content in the public space, they must be able to take responsibility for the consequences that may arise.

Resistance must be added to this duo. As said by Matthias Rath (2017, p. 8570):

> Until now resistance has been a reaction. We develop resistance in reaction to an unfulfilled responsibility – for example, by the media maker and the media vendor. However, as a media practice, resistance is something more. It must be understood as a quality in its own right, not merely an instrumental value for the defense of another. Resistance, instead, is the practice adopted by a media actor interpreting their responsibility for themselves. Offering fundamental resistance against the media presence of others in one's own lifeworld, as conditioned by an awareness of mediatization, lives up to responsibility. Transformed by the media, classical appetitive ethics that have traditionally asked 'what is good for me?' now come into their own once more: resistance is governed by the ability of a media actor to gauge the limits of accountability for their media practice. Since every communication conveyed by the media, every media practice, always involves other media actors. This is so either because individuals created the media offerings and made them publicly accessible to other users via the media, or because people use media content, offerings, and functions that were made publicly accessible by other media actors.

Resistance thus becomes the crucial ethical tool for taking up the challenge of moving from knowledge to participation. It allows everybody to exercise citizenship, acting in an increasingly responsible way to help other media users in becoming data literate. This is an example of the political vocation, in the highest sense of the term, that media literacy education has always been driven by and cannot neglect today.

References

Battista, A., & Conte, J. (2016). Teaching with data: Visualization and information as a critical process. In N. Pagowsky, & K. McElroy (Eds.), *Critical Library Pedagogy Handbook. Vol 2: Lesson plans* (pp. 147-154). American Library Association.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society, 15*(5), 662-679.

Brunton, F., & Nissenbaum, H. (2015). *Obfuscation: A User's Guide for Privacy and Protest*. MIT Press.

Buckingham, D. (2019). *The Media Education Manifesto*. Polity Press.

Carmi, E., Yates, S. J., Lockley, E., & Pawluczuk, A. (2020). Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review, 9*(2), 1-22.

Carrington, V. (2018). The changing landscape of literacies: Big data and algorithms. *Digital Culture & Education, 10*(1), 67-76.

Colombo, F. (2020). *Ecologia dei media. Manifesto per una comunicazione gentile*. Vita e Pensiero.

D'Ignazio, C., & Bhargava, R. (2015). Approaches to building big data literacy. In *Proceedings of the Bloomberg data for good exchange conference*.

Eugeni, R. (2021). *Capitale algoritmico. Cinque dispositivi postmediali (più uno)*. Scholé.

Floridi, L. (2014). *The Fourth Revolution. How the Infosphere is Reshaping Human Reality*. Oxford University Press.

Kuhn, T. (1962). *The Structure of Scientific Revolutions*. The University of Chicago Press.

Masterman, L. (1985). *Teaching the Media*. Routledge.

Missika, J.L. (1983). *La fin de la télévision* [The end of television]. Gallimard.

Pangrazio, L., & Selwyn, N. (2019). 'Personal data literacies': A critical literacies approach to enhancing understandings of personal digital data. *New Media & Society, 21*(2), 419-437.

Pappas, E., Emmelhainz, C., & Seale, M. (2016). Thinking through visualizations: Critical data literacy using remittances. In N. Pagowsky, & K. McElroy (Eds.), *Critical Library Pedagogy Handbook. Vol 2: Lesson plans* (pp. 179-187). American Library Association.

Pawluczuk, A., Yates, S., Carmi, E., Lockley, E., & Wessels, B. (2020). Data citizenship framework: Exploring citizens' data literacy through data thinking, data doing and data participation. *Digital Skills Insights*, 60-70.

Rath, M. (2017). Media Change and Media Literacy – Ethical Implications Of Media Education. In L. Gómez Chova, A. López Martínez, & I. Candel Torres (Eds.), *ICERI2017 Proceedings* (pp. 8565-8571). IATED Academy.

Rivoltella, P.C. (2020). *Nuovi alfabeti*. Scholé.

Špiranec, S., Kos, D., & George, M. (2019). Searching for critical dimensions in data literacy. In *Proceedings of CoLIS, the Tenth International Conference on Conceptions of Library and Information Science. Information Research, 24*(4), paper colis1922.

Thompson, J.B. (1995). *Media and Modernity: A Social Theory of the Media*. Polity Press.

van Dijck, J., Poell, T., & de Waal, M. (2018). *The Platform Society: Public values in a Connective World*. Oxford University Press.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Public Affairs.

3. Towards Dubai 2020 *Connecting Minds, Creating the Future*

Education and 'Artificial Intelligence'

P. Malavasi, T. Giovanazzi

ABSTRACT

What has human education became compared to the pressing digitalization of daily life? The close relationship that people have with technological devices equipped with advanced functions requires a new and deep, multi-level and transdisciplinary reflection. Dealing with great potential and inevitable risks, ethical-educational responsibilities and pedagogical guidelines are relevant.
The relationship between human education and radical technologies must be based on a critical view, a policy and a proper plan aiming the care for the common home. The emblematic issues of technological innovation and sustainability also represent the theme of the World Expo in Dubai (1 October 2021 - 31 March 2022). Connecting Minds, Creating the Future: an 'authentic' culture of scientific and technological research is based on the educational processes and creative dynamism that identify innovation as the space of experience for generating horizons of action for a sustainable and integral development of humanity.

The Ambiguity and Power of Technological Civilization: Education and 'Artificial Intelligence', Emblematic Issues

"We are both intrigued and frightened by the prospect of machines that can respond to us as a person would and, on a certain level, might even seem human" (Brooks, 2017, p. 52). It is within the framework of the history of human development that this volume aims to situate the relationship between civilization and instrumental devices, between the mystery of the person and the efficiency of machines.

Robotics was created in the middle of the last century as a sector of cybernetics and developed strong processes of integration with engineering research and established itself through the design of machines equipped with increasingly 'autonomous' information, assessment and action systems. The growing media interest in robots, often character-

---

Conceived as a single essay, this paper was written by Pierluigi Malavasi (first and second parts) and Teresa Giovanazzi (third and fourth parts).

ized by concerns and worries, is matched by their massive use in many industrial fields. The profound anthropological changes we have seen in the contemporary world, such as research in genetics and neuroscience, raise unprecedented questions (Ravasi, 2017). To dilute the excessive spread of alarmism, I share the opinion of Rodney Brooks (2017, pp. 59-61), long-time director of the Massachusetts Institute of Technology's Laboratory for Computer Science and Artificial Intelligence, on the topic: *Is artificial intelligence destined to dominate our lives?*

> There's a lot of hysteria in the debate about the future. Many people wonder how powerful robots will become, when it will happen, and what will happen to our jobs. [...] Almost every innovation in the field of robotics and artificial intelligence takes much longer to spread than predicted by the experts and external observers. [...] We won't be caught by surprise, I'm not saying there will never be problems, I'm saying they won't be sudden and unexpected[2].

The most popular representations of the so-called intelligent systems include the influential mythology of metallic automatons taking over from humankind or robots replacing their work activities and developing malevolent or violent attitudes. These stereotypes have been nurtured since the 1950s by science fiction, literature and filmography, which expressed a reaction to the astonishing scientific and technological advances and the continuing risk of nuclear catastrophe.

Open thinking and multilateral reflection on innovation must face major challenges, while avoiding carelessness and superficiality. Several issues are at stake. Among others, it is worth mentioning the very ambiguity of words such as *artificial intelligence* (Accoto, 2017) and *robotics*; the media's overestimation of the short-term effects of technologies; the real skills of robots and the likely improvement of their performance; the expectation or concern about the timing of deployment and employment repercussions; the popularity of science fiction mythology and, at the same time, the need for technological literacy/education.

The expression *educating robots* concerns industrial and service organization in the contemporary world and alludes to the need to include, among the purposes of human education, activities aimed at the conception, production and operational management of machines, supports and applications. First of all, it is a matter of becoming aware of what Adam Greenfield (2017) defines *radical technologies* or the design of

[2] The expression 'artificial intelligence,' coined in the 1950s, refers to the foundations, methodologies and techniques behind the functioning of computational machines; hardware and software systems able to perform certain cognitive tasks typical of natural intelligence. This is a name that is paradoxical in some ways, certainly ambiguous and naturally polysemic.

everyday life in contact with tools that are equipped with advanced functions and, paradoxically, 'intelligent'. The smartphone is the ultimate symbol here, documenting, in a protean way, the modern human relationship with a machine that has become almost indispensable for organizing our daily existence.

Very few objects have been so omnipresent and pervasive in the history of civilizations (Ferraris, 2005). For many of us, the smartphone is the last thing we pay attention to before we fall asleep and the first thing we pick up when we wake up. We use it to meet people, to communicate, to have fun, to orient ourselves, to buy and to sell. However, we also use it to document the places we go to, the things we do, our acquaintances; to fill the empty spaces, the moments of pause and the silences that used to occupy part of our existence.

> Smartphones have effectively altered the entire fabric of daily life, by completely reorganizing longstanding spaces and rituals, and transforming others beyond recognition. At this historical juncture, it is simply impossible to understand the ways through which we know and practice the world around us without having at least some idea of how the smartphone works and the various infrastructures on which it depends. Its ubiquity, nevertheless, makes it an object that is anything but trivial. We use it so often that we fail to clearly understand what it really is; it has appeared in our lives so suddenly, and in such an all-encompassing way, that the scale and force of the changes it has brought about largely elude our awareness (Greenfield, 2017, p. 7).

Although, strictly defined, a smartphone is not a robot, it is difficult to negate that it represents a formidable emblem of the performative relevance of technology. Using networked digital information, it has become the dominant way in which we experience everyday life. In educating people to use smartphones responsibly, a window of sense is opened on the web of technical, financial, legal and operational connections and agreements that constitute not only a technological device but an ecosystem of services. Equally important is pedagogical analysis and ethical-educational discernment regarding the complexity of issues posed by the planetary network of perceptions and responses defined as the *Internet of Things*: here, computability and data communication are embedded and distributed within our environment, in its entirety (Kuniavsky, 2010).

The *Internet of Things*, in many aspects analogous to the smartphone, is an *assemblage* of technologies, perception *regimes* and operating protocols, identified with an umbrella term familiar to a large part of the public. What unites very heterogeneous elements is a conception of the world that connects devices, applications, supplier companies and the performance of products and services in order to make everyday life sit-

uations sensitive for the network and available for analysis and processing.

> Although this colonization of the quotidian may be perceived as something that develops autonomously, without manifest guidance or other urgent justification than the fact that it is our technology that makes it possible, it is always better to keep in mind how certain ambitions come into play. Some of these relate to commercial needs [...]. Others are based on a set of interests that may relate to the management of the infrastructure that secures public utility goods. Inevitably, some of these ambitions involve surveillance, security, and control (Greenfield, 2017, pp. 32-33).

We are heirs to two centuries of progress that express the extraordinary creativity of humankind. Nanotechnology, robotics, biotechnology make the availability of products and services to improve quality of life possible. However, it cannot be ignored that never before have knowledge and power, especially economic power, offered such effective tools for the manipulation of consciences and the domination of the whole world. There is a tendency to believe that any acquisition of power is simply an increase in well-being and life force. Our age tends to develop poor self-awareness of its own limits: "The possibility of misusing power is constantly increasing when there are no norms of freedom, but only claims to necessity, utility and security" (Guardini, 1965, p. 87). And nothing ensures that humanity will make fair and supportive use of the digital revolution, big data, and robotics in the future[3].

> The immense technological growth has not been accompanied by a corresponding development of the human being in terms of responsibility, values, conscience [...] as shown by the atomic bombs launched in the middle of the twentieth century, the great deployment of technology flaunted by Nazism, Communism and other totalitarian regimes in the service of the extermination of millions of people, without forgetting that today's wars employ increasingly deadly instruments. In whose hands is so much power and into whose hands can it fall? It is terribly risky that this power lies at the fingertips of a small section of humanity (Francesco, 2015, pp. 104-105).

The need for an alliance between robotics and pedagogy is evident from the facts: a historical analysis of the use of technology shows the need for a constant relationship between the discourse on education

[3] In the variety of use of big data and robotics, however, it is necessary to recognize their ethical relevance in reference to the medical-health sector, as regards data management systems and the personalization of care through predictive models aimed at pursuing the increase in the efficacy of therapies. In this regard, see Valentini, Dinapoli, and Damiani (2013).

and the development of technology, and between training on how the use of tools and ethical-moral consciousness. Our close relationship with machines endowed with advanced functions operating in concrete ways in daily life points to a new and deep awareness of the human potential that gives rise to intelligent devices – such as the smartphone – and to real robots. The responsibility for accessing and using sensitive information, for the most diverse of uses, now affects anyone using the internet.

The notions *algorithm* and *computational thinking*[4] are becoming more and more widespread in the discourse on the frontiers of education. Among the main reasons cited on the pages of this volume is recognition of the widening of the educational sphere that the pressing digitalization of everyday life entails. In a specific way, the task of thinking about the relationship between human and machine learning, before assuming didactic and operational relevance, is of fundamental importance in the field of pedagogical reflection. The relationship, the alliance between robotics and pedagogy, with its latent ambiguities and legitimate suspicions, is an emblematic theme.

A further goal and a new challenge for humans has emerged, the development of which – through *robotics*, *big data* and the *Internet of things* – brings with it the need for a critical and conscious approach to machines, a policy, an educational plan for the care of our common home.

The Technocratic Paradigm and New Humanism: Awareness, Education and Responsibility

We decide to give meaning and sentimental value to machines (if we are capable of becoming attached to our old car, it is even more likely that we can love a human-like robot by projecting our ancestral parental instincts onto it). In the same way, we are the ones who attribute to the machine the ability to see us as human beings and to establish a relationship with us. In reality, the machine has no relationship with us in the human sense of the term (Cingolani & Metta, 2015, pp. 7-8)[5].

---

[4] The widespread dispersal of primary literacy courses and of recreational-educational activities concerning both computer programming languages (coding) and so-called 'educational robotics' aimed at children, attests to the need for a well-structured scientific-cultural and, specifically, pedagogical approach.

[5] "Robots are not necessarily anthropomorphic; indeed, in most cases they are far from having any human semblance" (Cingolani & Metta, 2015, pp. 7-8). However, it is the humanoid robot on which it becomes very easy to project concerns and expectations, because of the likeness and functioning that must relate to humans.

It *interacts*, through complex artificial intelligence algorithms, talks, drives a car, and makes small operational decisions, but it does not feel emotions, nor does it have any sentimental structure or personal status.

However, sophisticated *technical products* – such as smartphones or humanoid robots – are capable of *mediating* patterns of relationships and directing the interests of well-identified power groups. Economic choices that appear to be simply instrumental are actually intentionally related to the type of social life they are intended to promote. It is inconceivable nowadays to think of technology as something purely functional; "The technocratic paradigm has become so dominant that it is very difficult to disregard its resources, and even more difficult to use its resources without being dominated by its logic" (Francesco, 2015, pp. 107-108). It tends to exercise hegemonic dominance over politics and economics, and the latter is willing to assume any technological progress in the key to profit maximization, without paying attention to any negative consequences for human beings. For many reasons, it is not a question of making refined distinctions between different economic theories, but rather to recognize the actual centrality of a reductive and individualistic mainstream in the factual orientation of the economy. Researchers and managers who do not admit this clearly nevertheless support it inadvertently through their failure to systematically address issues such as the dignity of work, the growing inequality in the distribution of wealth and the continuing ecological crisis in relation to the rights of future generations. Through their behavior they affirm that the main objective corresponds to the marginal increase in profits. "The market alone, however, does not guarantee integral human development and social inclusion" (Benedetto XVI, 2009, n. 35). "We have not identified with sufficient clarity the deepest roots of today's imbalances, which are related to the orientation, goals, meaning and social context of technological and economic growth" (Francesco, 2015, nn. 108-109).

It is not the power of technology that is the fundamental issue, it is the "loyalty of the person to other human beings: the loyalty, responsibility, and respect that establish the educational relationship between people" (Colicchi Lapresa, 1994, p. 110). It is necessary to be aware of a disorientation perhaps even greater than the one described by Marshall McLuhan in 1964, regarding the rapid rise of electricity: "The technique of electricity is in our midst and we are stunned, deaf, blind and dumb in the face of its collision with the technique of Gutenberg" (1964, p. 18). Through the digital revolution, technological pervasiveness is greatly increased. Byung-Chul Han (2015, p. 9) notes how "we are reprogrammed, without fully understanding this radical paradigm shift. We are behind the digital *medium* which, acting below the level of conscious decision, decisively modifies our behavior, our perception, our sensitivi-

ty, our thinking, our living together. Today we become intoxicated with the digital *medium*, without being able to fully assess the consequences of such intoxication".

This blindness to the implications of the changes and the simultaneous daze represent an irreducible component of the crisis of civilization, which is evolving into a great cultural, spiritual, and educational emergency. It implies "long processes of regeneration" (Francesco, 2015, n. 202), affecting every level of economic, political and social life, as well as daily life itself.

The need for a *new humanism* arises from a new organization of the *everyday life experience*, which can be considered fundamentally *narrative*. Individuals, both narrating and narrated, and the interpersonal relationships they weave are not exhausted on the plane of logical-formal knowledge and technical application; they are always characterized by a dynamism of action and interpret *existence in the flesh*. "The desert grows in amplitude because at the same pace as the physical-geological-geographical desert, the desert that everyone hides within themselves grows to a greater extent and at a higher speed, that is, the aridity of the soul, of the heart and even of the mind that leads to pursue its own short-term profit at the expense of others, of contemporaries" (Anelli, 2016, p. 10), and of those to come.

Among technocratic paradigms and *humanoid robots*[6], we need a development that is marked by "that extraordinary valorization of individuals, regardless of their age and other concrete determinations, accompanied, however, by an equally important valorization of what is outside the individual: the natural and social world" (Bertolini, 1994, pp. 31-32). We must care for the world around us, and have respect for humans, throughout our lives.

"To *respect*, literally, means to *look away*. It is a regard. In relating respectfully to others one refrains from pointing one's gaze indiscreetly. Respect presupposes a detached gaze, a pathos of distance. Today, this gaze yields to a vision devoid of distance, which is typical of entertainment" (Han, 2015, p. 11). A society without respect, without *pathos* of

[6] Note in Cingolani and Metta (2015, pp. 7, 167 and 176): "The humanoid robot lends itself to become the surrogate of the human, with its great mysteries regarding creation, death, feelings [...] Humanoid robots can be classified based on parameters including size, possible use (research, leisure, home care, industry, natural and environmental disasters) and their cost. [...] iCub is the humanoid robot, designed by the Italian Institute of Technology in Genoa (IIT), which has become the most widely used humanoid platform in the world. An open-source strategy has led to its deployment in leading robotics research centers in several countries, including Japan and South Korea. The robot has 53 joints and a human-like sensory system that includes cameras, microphones and inertial, strength and touch sensors".

distance results in sensationalism, disinterest in depth and in indifference.

"We are faced with a common responsibility towards all humanity, especially towards the poor and future generations" (Benedetto XVI, 2010, n. 2). "Robot design is part of the perspective of an integral ecology made up of community networks. It is through many simple daily gestures that we contribute to breaking the logic of violence, exploitation" (Francesco, 2015, n. 230), and greed. Cingolani and Metta (2015, p. 45) note in this regard:

> There is a problem of education and social awareness, which not only serves to steer human behavior in the right direction, but also to maintain a high level of attention to the dangers that are not intrinsic to a technology per se, but which can result from its misuse or unexpected effects. To give an example, no one would have ever imagined that a civil aircraft could be a weapon of mass destruction, but the tragic events of September 11, 2001, have shown that the irresponsible use of any technology makes it dangerous. [...] Education in the proper use of technologies presupposes an ethical culture, as well as a scientific-technological one, without which humanity is not able to manage the results of its knowledge.

In his *Address to participants at the plenary assembly of the Pontifical Academy for Life 2019*, Pope Francesco observes,

> A global bioethics is an important front on which to engage. It expresses an awareness of the profound impact of environmental and social factors on health and life. It is an approach very much in tune with integral ecology, described and promoted in the Encyclical *Laudato si'*. [...] The possibility of acting on living matter at ever smaller orders of magnitude, of processing ever larger volumes of information, of monitoring – and manipulating – the cerebral processes of cognitive and deliberative activity, has enormous implications: it touches the very threshold of the biological specificity and spiritual difference of the human being. *The difference of human life is an absolute good*[7].

The topic of 'emerging and converging' technologies must be made the subject of programmatic and incisive attention by pedagogy, in dialogue with the *hard sciences* and the humanities. Artificial intelligence, big data, robotics, and technological innovations in general must be employed in ways that contribute to the richness of human education, the service of populations, and the protection of our common home, rather than

[7] Address of the Holy Father Francesco to participants at the plenary assembly of the Pontifical Academy for Life. See https://www.vatican.va/content/francesco/it/speeches/2019/february/documents/papa-francesco_20190225_plenaria-accademia-vita.html.

the exact opposite. The inherent dignity of every human being must be placed tenaciously at the center of reflection and action.

In this sense, it is useful to recognize that the denomination *pedagogy of artificial intelligence*, although certainly effective to mark the challenges of a pedagogy aimed at reflecting radically on technologies, is ambiguous and may lead to misunderstandings. The terms should not conceal the fact that the functional automatisms of a machine are far removed from the human prerogatives of knowledge and action, consciousness and intentionality. We need an *authentically sustainable development*, not evasion from the commitment to a *pedagogy of artificial intelligence*, in the face of responsibility for future generations.

Universal Exhibition, Educational Values, Responsibility

In today's social and cultural scenario characterized by accelerated technological 'progress', this contribution touches on some elements that I consider useful for addressing emblematic issues connected with innovation and sustainability, with particular reference to the theme *Connecting Minds, Creating the Future* of the next universal exhibition to be held in Dubai from 1 October 2021 to 31 March 2022[8], postponed by one year due to the COVID-19 pandemic. What value does the universal exhibition assume in the field of pedagogical research? First of all, it is necessary to reflect on the importance that the events have marked throughout history, to grasp the meaning that today can be attributed to the universal exhibition.

The advent of mechanical industry favored the birth of large exhibitions and, although it was England in 1851 that first hosted the *Great Exhibition of the Works of Industry of All Nations*, the concept was born in France at the end of the XVIII century. In fact, in 1798 the *Exposition Publique des Produits de Industrie Française* was held in Paris, in the Champ-de-Mars, for the explicit purpose of promoting ideas and values aimed at progress and a renewed feeling of national identity (Crippa & Zanottera, 2008). The pressure of industrial innovations required a space and a time frame in which to focus attention on scientific and technological research, showing progress and discoveries, in societies where the means of communication and the exchange of news were often patchy and occurred at a very relaxed pace. This enabled mass access to the widest range of information possible with the aim of stimulating the comparison of ideas and their gradual overcoming, considerably increasing the development already under way.

[8] See https://www.expo2020dubai.com/.

The universal exhibitions were opportunities to give voice to different nations: they were international exhibition events of a non-commercial nature lasting more than three weeks, officially organized by one nation and in which other countries were invited to participate through diplomatic channels. The subject of these exhibits were universal themes that affected the full range of human experience; each of them focused attention on a particular theme, on which each participating nation could express its opinion within a pavilion specially created or provided by the host city (Dell'Osso, 2008).

They symbolized a moment of trust, an extraordinary opportunity to improve conditions of life on the planet: through them, humanity represented itself and reached out to the future. Expressions of the society and industrial culture of the time, the events contributed to the improvement of construction techniques in times of scarcity and played important roles in the cities that hosted them, giving impetus to urbanization, the identification of axes of urban growth, enrichment and the provision of infrastructures and services for the city. In this context, pedagogical research is called upon to face the challenge that arises between the "realization of events increasingly oriented towards immediate economic profit and therefore aimed at the success of the event in itself, on the one hand, and exhibitions that are planned with the intention of producing shared and protracted benefits, respecting the real needs of the host city and with reference to the development process that the theme of the event hopes and prefigures" (Malavasi, 2013, p. 101). The great popularity of the initial events, the influx of visitors and the notoriety they generated, meant that from the end of the XIX century onwards the number of events and their frequency increased considerably. Consequently, the need arose to establish guidelines to prevent the uncontrolled proliferation of exhibitions and to provide greater guarantees to participating countries.

The official institution that still approves, regulates and controls exhibitions today is the *Bureau International des Expositions* (BIE), established in Paris on November 22, 1928, with an International Diplomatic Convention ratified by 31 founding nations including Italy (currently 170 members)[9], that establishes the rights and responsibilities of the organizers of the exhibitions and of the participating bodies. The action of the *Bureau International des Expositions* is based on three fundamental guiding principles: trust, solidarity and progress. According to what is stated in the first article of the Institutive Constitution of the *Bureau Internation-*

[9] The regulation came into force in 1931, Barcelona (1929) was the last exhibition organized according to the previous parameters, with Chicago (1933) the first to follow the new ones.

al des Expositions, each event, regardless of its name, has a mainly educational purpose towards the public, by recognizing the means available to human beings to satisfy the needs of civilization and make the prospects for the future emerge from one or more sectors of human activity, supporting innovation as an instrument at the service of human progress[10].

The exhibitions of the third millennium have undergone a progressive paradigm shift, becoming unique opportunities to reflect on the responsibility of education in the era of sustainable development, a chance to come together and debate, on a global scale, issues of global interest. The category of responsibility arises as a key element of a relationship in solidarity with life on Earth: it represents both a fundamental condition for the training needs that arise from the construction of civil society, and an awareness linked to the ultimate truth of the human being, in the face of a world in constant evolution and change. The current ecological crisis, in various respects, has the character of a moral issue with significant social implications and urgently calls for pedagogical reflection on educational possibilities, critically engaging the principles of effectiveness and efficiency in a dialectic between local and global. This issue involves all sectors of human life, calling into question the pedagogical research on the environment to develop new ideas and projects and to adopt the vision of the future as an open space, a place for continual exchanges of knowledge and comprehension, between human education and the protection of creation from an axiological perspective, for the protection of life itself in its various historical-cultural forms (Iavarone et al., 2017).

In the 2000 Hanover Expo *Man, Nature, Technology*, the concept of nature and technology understood in the broadest dimension of human experience were given considerable prominence. The prospect of finite resources and the awareness of environmental damage caused by the exploitation of the ecosystem became increasingly relevant from the Seventies onwards, sensitizing those who work in the field of architecture. An interpretation that for the first time linked the concept of progress to that of resources, renewal and environmental sustainability. An in-depth reflection was developed on natural and eco-sustainable materials such as wood, to be used in construction, on the need for recycling and on the relationship with the environment. The positive signal conveyed by the event was an awareness that sustainable development and architectural quality could equate to the experimental use of materials that have always been used in construction techniques. Despite the relevance of the theme, the Hanover exhibition took a largely commercial

[10] See https://www.bie-paris.org.

and purely entertainment-driven slant, to the detriment of the educational content (Beltrame, 2014).

In 2010, the Shanghai international exhibition *A Better City, a Better Life*, put the metropolis with all its intrinsic potentials at the center of the investigation. The exhibition promoted a fruitful discussion on the issues of urban planning, impact assessment and the relaunch of regional quality within the global arena. The debate involved the analysis of social systems and local realities, noting critical issues and formulating proposals for redeveloping the urban fabric, with reference to the dynamics of Chinese growth. More than a theoretical, educational reflection on the ways and forms to give life to a new civilization, the Shanghai Expo constituted a political showcase for the host country with its strong desire to acknowledge its place as a leading country on a global level.

The last event before Dubai 2020 was held in Milan in 2015 on the theme *Feeding the Planet, Energy for Life,* taking a considerably innovative approach to nutrition and proposing a new model of human development for the XXI century. It has qualified as a world event with a strong cultural connotation, a 'laboratory' of ideas to explore and experiment with the theme of a healthy, safe and adequate diet for all peoples, whilst respecting the planet and its balance: a research and development center in which to profile the value of human beings in their relationship with their regions, traditions and future (Giovanazzi, 2018). The event was characterized as a unique opportunity for sustainable development for the present, but also for the immediate future, as the actions of each individual have significant repercussions on the community and the living environment (Di Vita, 2010). The dimension of sustainability has been confirmed, as an inspiring principle, through the use of new materials and innovative techniques in the construction of the buildings designed and in their management. The exhibition container represented an initial form of meaning, a harmonious relationship between human beings and nature by guaranteeing the conditions of environmental sustainability, transmitting an educational message aimed at interrogating the minds of observers in order to solicit reflection and find solutions to serious food issues.

While Milan in 2015 posed the question of food safety as a factor of growth and development, Dubai will propose the theme of sustainability that requires connections and connectivity to design a future in which health-related solutions are firmly integrated with environmental ones; the fight against climate change with the exploitation of natural resources; and the growth of new digital skills for social and economic development. Sustainability introduces the dimension of the future and highlights the irreversibility of human action, promoting the potential of each of us to nourish trust in the possibility of transforming reali-

ty and making progress fair and widespread under the banner of a new planetary civilization. This involves leading the design of training interventions "to make solidarity ties sprout, generative alliances of authentic development" (Vischi, 2020, p. 14) to create innovative, sustainable and inclusive communities that recognize and protect the value of the environment.

Towards Dubai 2020 "Connecting Minds, Creating Future": Between Innovation and Sustainability

What emblematic meaning does pedagogy, a critical discourse on educational experience and training processes, assume in the era of the technical reproducibility of the world and the globalization of markets?

> What is at stake is not only the cost/benefit ratio of the actions carried out by individuals in social and natural contexts, but, in a radical way, the centrality of people in realizing the present and future of life on earth, in conceiving their right to a relationship in harmony with nature; the moral responsibility to provide an integral education in the circle of creation, in the face of possible catastrophe, is under discussion (Malavasi, 2017, p. 19).

The challenge is to combine education, environmental protection and economic-technological practices of social life and world views, so that the pedagogical discourse exercises its role of "practical design knowledge" (Elia, 2014, p. 26) with regard to human education in the confrontation between culture and civilization, between purposes and values. The knowledge that is processed and made available has to be aimed at increasing the desire to nurture and support innovation in the name of the future, of a prosperous future on the horizon of sustainability. The awareness of ecological problems, through a multidisciplinary approach, requires a new ability to analyze the relationships between civil society, institutions and the business world according to a multiplicity of criteria and methods, to identify strategies and intervention tools, combining developments in sustainable science and technology with environmental protection.

The universal exhibition in Dubai 2020 *Connecting Minds, Creating Future* will be held in the ME.NA.SA. areas in order to fit into the reference framework (Middle East, North Africa, South Asia), and is the first to take place in an Arab country. It aims to connect the thinking skills of the whole world, mobilizing everyone to engage in our shared challenges. Focusing on unprecedented development and innovation, it invites to think about the progress of civilization according to a model that is in-

tegrated and, in some respects, in the future, 'cooperative', in terms of the relationship between human and machine. Of considerable importance, in this reflection, is the *Humane Technology Lab*, a multidisciplinary laboratory of Università Cattolica del Sacro Cuore, which questions the processes by which changes are taking place in the local and global contexts and, specifically, in the relationships between technologies and different dimensions of human experience, on the psycho-social, pedagogical, economic, juridical and philosophical levels[11]. What is the impact of technology on daily life? How do emerging technologies, particularly robotics, artificial intelligence and virtual reality, affect our life today? The spheres of technology and life intersect and intertwine, to the point of "cross-pollinating each other in an unprecedented and original way, producing transformations with respect to which pedagogy needs to ask questions, develop analyses and perspectives, new directions of intervention" (Pinto Minerva & Gallelli, 2004, p. 17).

The theme of innovation is called upon to define new processes and different models capable of responding to vital issues raised by contemporary society, from the protection of creation to the quality of life of all people. It is characterized by risk taking, being related to creative invention and allowing us to establish a continuum in the process that solicits change. This means defining objectives and implementation methods, interdisciplinary training courses and involvement and verification tools with a view to systematically addressing the problems associated with the impact of production, distribution and consumption activities and their eco-sustainable conversion, whose technical and management aspects are strictly connected with the ethical-educational and political-economic ones.

Making use of technology, oriented and placed at the service of a "healthier, more human, more social and integral type of progress" (Francesco, 2015, p. 112) could be a plausible keystone for a harmonious future life on the planet, posing new questions about human responsibilities. Pedagogical planning as a discourse aimed at safeguarding the environment, the common good of humanity, invites us to "exercise responsible government over nature in order to safeguard it, exploit it and cultivate it in new forms and with advanced technologies in ways that it can suitably welcome and feed the people who live there" (Benedetto XVI, 2009, n. 50).

The reflection of some thinkers including Hans Jonas (1979), who

[11] The *Humane Technology Lab* (HTLab) of Università Cattolica del Sacro Cuore was created with the aim of investigating the relationship between human experience and technology, promoting research activities both in the academic field and from the perspective of cultural dissemination to a wider audience. See http://www.humanetechnology.eu/.

elaborated an ethics for the technological age according to the principle of responsibility, is emblematic: an invitation to act in such a way that the consequences of one's actions are compatible with the continuation of an authentically human life on Earth. Technological innovation imposes a new dimension of responsibility on ethics in endorsing a culture of progress, aimed at promoting and cultivating values capable of realizing processes of real humanization while respecting the environment. "Educational responsibility and the formation of conscience and judgment in the face of artificial intelligence represent research horizons whose relevance must be measured with the concreteness of applications and lines of technological development" (Malavasi, 2019, p. 101). The discourse on education aims to critically define, in terms of planning, the interaction between the human community and technological development, and interprets the multiple cultural elements that intervene to determine a certain technocratic anthropology.

The center of the Universal Exposition site will be Al Wasl Square, named after Dubai's ancient name, meaning 'the connection'. From it will unravel, like three large 'petals', the thematic areas of the event that will explore and inspire the intent to 'create the future': *sustainability* – progress that does not compromise the life and needs of the next generations, ensuring accessibility and resilience of environmental, energy and water resources; *mobility* – the creation of new and more efficient physical and virtual connections between people, communities and countries by means of innovative logistics, transport and communication systems to transform the way we live and exchange knowledge and ideas; and *opportunity* – unlocking the inner potential of individuals and communities and being an agent of change for a better future.

Among the thematic areas there will be three important exhibition structures: the Welcome Pavilion, the Innovation Pavilion and the United Arab Emirates Pavilion, with the national pavilions built outside of these. In Dubai, the model of exhibitions changes: moving from the principle of national identity to one of relationship, the exhibition becomes a sort of platform to which each participating country will bring its own ability to connect and interact with others, through generative networks characterized by cultural proximity and encounter[12]. The dimension of sustainability underlying the universal exhibition is a condition and object of human education and economic wealth for developing goods and services in which the interpersonal relationship becomes a resource, aimed at addressing the irresponsible ways Earth's assets are being managed, and bringing prosperity to every single person. It follows the construction of a coexistence between countries in close con-

[12] See https://www.expo2020dubai.com/.

nection with the values of the educating society, recognizing the ecological question as a challenge to which the global community is called upon to respond. Promoting dialogue on the most appropriate ways to build the future of the planet draws on humanity's ability to collaborate, 'connect minds', and develop the vision of a way of life that is conscious, respectful and jointly responsible for the needs of all people. This environmental challenge concerns all of us and requires us to come together in a new form of universal solidarity to protect creation, whilst each maintaining our own culture, experiences and values.

For six months Dubai will be transformed into a world showcase in which more than 190 participating countries will present the world with their ideas, projects, exemplary and innovative models in the field of tangible and intangible infrastructures on the themes of the universal exhibition, to promote new forms of knowledge by embracing unprecedented and significant change for the global well-being of humanity. The objective of the event is to create a sustainable ecosystem, recyclable materials and renewable energy components, demonstrating the progress made by technology aimed at improving quality of life. The model of sustainable civilization that we are called upon to build, in order to interpret the relationship between the global context and strengthen the ties with the surrounding environment, will have to create effective correspondence between fundamental human values, the lowest environmental impact and technological innovation.

All progress in knowledge and technology makes human wisdom more necessary, rediscovering the inescapable perspective of reference for the enhancement of the past, the management of the present and the creation of the future. This is significant in directing the educational process towards those horizons of humanity which connote the specific nature of human nature. "Man is the most precious asset of the entire creation, and it is only from a discourse on man that he can develop an ecological reflection that aims to solve problems rather than to feed sterile debates or poetic proclamations" (Anelli, 2016, p. 7).

Radical change is necessary and requires both training policies and institutional interventions in order to generate green sensitivities and ethical guidelines to find unprecedented ways of thinking about the surrounding world. There is an awareness of identifying and following new paths of human and economic development that are in balance with natural systems. It follows the essential need to redefine progress understood to mean a better life for all, without exclusion: "Economy and progress must be rethought in the perspective of favoring integral development and, for this reason, the human sense of ecology must be rediscovered" (Magnoni, 2015, p. 20). A more conscious ecological view makes it possible to promote proper management of natural resources, recog-

nizing that the relative supporting technological progress offers a means of reducing their exploitation and transformation, with the relative decrease in the capacity for human action on nature.

Connecting Minds, Creating Future implies enhancing a culture of scientific and technological research oriented towards spaces for critical reflection on training processes, identifying in innovation the possibility of opening new horizons of action for sustainable and integral development. Beyond a merely instrumental use of radical technological innovation (Greenfield, 2017), through a critical look at the conception, production and operational management of machines equipped with advanced functions, this kind of innovation can contribute to the ethical-transformative force of pedagogical reflection, open to multidisciplinary dialogue, starting from the importance of education in addressing environmental issues within a culture of sustainable living.

In light of the dynamism of transformation and the pervasiveness of the digitalization of the world, thinking about a *pedagogy of artificial intelligence* is an educational challenge for the development of civilizations, based on a vision of humanism structured around its engagement with the potential and limits of technological processes. The approach of pedagogical reflection to the theme of the next universal exhibition embraces the possibility of attributing an educational sense to technology as a wealth of human formation, used to benefit peoples and to safeguard our shared home. *Person*, *environment* and *technological innovation* are closely interconnected and involve the search for values and principles to educate us on a sustainable way of life that promotes the fair and supportive development of humanity.

References

Accotto, C. (2017). *Il mondo dato. Cinque brevi lezioni di filosofia digitale*. EGEA.

Address of the Holy Father Francis to participants at the plenary assembly of the Pontifical Academy for Life. In https://www.vatican.va/content/francesco/it/speeches/2019/february/documents/papa-francesco_20190225_plenaria-accademia-vita.html.

Anelli, F. (2016). La natura come creazione e le responsabilità dell'uomo. In C. Giuliodori, & P. Malavasi (Eds.), *Ecologia integrale. Laudato si'. Ricerca, formazione, conversione* (pp. 3-10). Vita e Pensiero.

Beltrame, M. (2014). *Expo Milano 2015. Storia delle esposizioni universali*. Meravigli.

Benedetto XVI (2009). Lettera Enciclica *Caritas in veritate*.

Benedetto XVI (2010). *Messaggio per la XII Giornata Mondiale della Pace - Se vuoi coltivare la pace, custodisci il creato*.

Bertolini, P. (1994). La mia posizione nei confronti del personalismo pedagogico. In G. Flores D'Arcais (Eds.), *Pedagogie personalistiche e/o pedagogie della persona* (pp. 31-54). La Scuola.

Brooks, R. (2017, November 24). L'intelligenza artificiale dominerà le nostre vite?. *Internazionale, 1232*, 52-61.

Bureau International Des Espositions. At https://www.bie-paris.org.

Cingolani, R., & Metta, G. (2015). *Umani e umanoidi. Vivere con i robot.* il Mulino.

Colicchi Lapresa, E. (1994). Persona e verità dell'educazione. In G. Flores D'Arcais (Ed.), *Pedagogie personalistiche e/o pedagogie della persona* (pp. 89-112). La Scuola.

Crippa, M. A., & Zanottera, F. (2008). *Expo x Expo. Comunicare la modernità. Le Esposizioni Universali 1851-2010.* Triennale Electa.

Dell'Osso, R. (Ed.) (2008). *EXPO da Londra 1851 a Shanghai 2010 verso Milano 2015.* Maggioli.

Di Vita, S. (2010). *Milano Expo 2015. Un'occasione di sviluppo sostenibile.* FrancoAngeli.

Elia, G. (2014). Il contributo della pedagogia come sapere pratico-progettuale. In G. Elia (Ed.), *Le sfide sociali dell'educazione* (pp. 26-46). FrancoAngeli.

EXPO Dubai 2020. At https://www.expo2020dubai.com/.

Ferraris, M. (2005). *Dove sei? Ontologia del telefonino.* Bompiani.

Francesco (2015). Lettera Enciclica *Laudato si' sulla cura della casa comune.*

Giovanazzi, T. (2018). *L'eredità educativa di Expo Milano 2015. Pedagogia dell'ambiente, alimentazione, ecologia integrale.* EDUCatt.

Greenfield, A. (2017). *Radical Technologies. The Design of Everyday Life.* Verso Books.

Guardini, R. (1965). *Das Erde das Neuzeit.* Grünewald.

Han, B. C. (2015). *Nello sciame. Visioni del digitale* [In the swarm. Visions of the digital]. Nottetempo.

Humane Technology Lab. At http://www.humanetechnology.eu/.

Iavarone, M. L., Malavasi, P., Orefice, P., & Pinto Minerva, F. (Eds.) (2017). *Pedagogia dell'ambiente 2017. Tra sviluppo umano e responsabilità sociale.* Pensa MultiMedia.

Jonas, H. (1979). *Il principio responsabilità. Un'etica per la civiltà tecnologica* [The responsibility principle. An ethics for technological civilization]. Einaudi.

Kuniavsky, M. (2010). *Smart Things: Ubiquitous Computing User Experience Design.* Elsevier.

Magnoni, W. (2015). Perché il cuore non si svuoti. In W. Magnoni & P. Malavasi (Eds.), *Laudato si'. Niente di questo mondo ci è indifferente. Le sfide dell'enciclica* (pp. 17-34). Centro Ambrosiano.

Malavasi, P. (2013). *Expo Education Milano 2015. La città fertile.* Vita e Pensiero.

Malavasi, P. (2017). Pedagogia dell'ambiente, educazione allo sviluppo sostenibile,

responsabilità sociale. In M. L. Iavarone, P. Malavasi, P. Orefice & F. Pinto Minerva (Eds.), *Pedagogia dell'ambiente 2017. Tra sviluppo umano e responsabilità sociale* (pp. 15-56). Pensa MultiMedia.

Malavasi, P. (2019). *Educare robot. Pedagogia dell'intelligenza artificiale*. Vita e Pensiero.

McLuhan, M. (1964). *Understanding Media: The Extension of Man*. Penguin.

Pinto Minerva, F., & Gallelli, R. (2004). *Pedagogia e post-umano. Ibridazioni identitarie e frontiere del possibile*. Carocci.

Ravasi, G. (2017). *Adamo, dove sei? Interrogativi antropologici contemporanei*. Vita e Pensiero.

Valentini, V., Dinapoli, N., & Damiani, A. (2013). The future of predictive models in radiation oncology: From extensive data mining to reliable modeling of the results. *Future Oncology, 9*(3), 311-313.

Vischi, A. (2020). Education for sustainable development, conversione ecologica, patto educativo. In A. Vischi (Ed.), *Global Compact on Education. La pace come cammino di speranza, dialogo, riconciliazione e conversione ecologica* (pp. 11-16). Pensa MultiMedia.

4. Positive Technology and COVID-19

G. Riva, F. Mantovani, B.K. Wiederhold

ABSTRACT

The past 10 years have seen the development and maturation of several digital technologies that can have a critical role to enhancement of happiness and psychological well-being. In particular, the past decade has seen the emergence of a new paradigm: 'Positive Technology', the scientific and applied approach to the use of technology for improving the quality of our personal experience. In this article we discussed the potential of Positive Technology to augment and enhance the existing strategies for generating psychological well-being during the COVID-19 pandemic. In particular different positive technologies – mHealth and smartphone apps, stand-alone and social virtual reality, video games, exergames, and social technologies – have the potential of enhancing different critical features of our personal experience – affective quality, engagement/actualization, and connectedness – that are challenged by the pandemic and its social and economic effects. In conclusion, although the focus of tackling the direct impact of COVID-19 is important, positive technologies can be extremely useful to reduce the psychological burden of the pandemic and to help individuals in flourishing even during difficult and complex times.

Introduction

The past 10 years have seen the development and maturation of several digital technologies that can have a critical role to enhancement of happiness and psychological well-being. In particular, the past decade has seen the emergence of a new paradigm: 'Positive Technology' (Gaggioli & Riva, 2014; Riva, 2012; Riva et al., 2012; Villani et al., 2016; Riva et al., 2019; Wiederhold & Riva, 2012) that can be described as the scientific

---

This chapter was originally published as Riva, G., Mantovani, F., & Wiederhold, B.K. (2020). Positive technology and COVID-19. *Cyberpsychology, Behavior, and Social Networking, 23*(9), 581-587. No competing financial interests exist. The preparation of this article was supported by the UCSC D3.2 2020 project "Behavioural change: prospettive per la stabilizzazione di comportamenti virtuosi verso la sostenibilità".

and applied approach to the use of technology for improving the quality of our personal experience. The foundations of this vision come from the 'Positive Psychology' approach (Fredrickson, 2001; Inghilleri et al., 2015; Seligman, 2004; Seligman & Csikszentmihalyi, 2000), a growing discipline whose main aim is to understand human strengths and virtues, and to elevate these strengths to allow individuals, groups, and societies to flourish (Gaggioli et al., 2016; Riva et al., 2015; Riva et al., 2016).

In these months the world is facing a complex crisis generated by the outbreak of a novel coronavirus-caused respiratory disease (COVID-19) (Huang et al., 2020). The impact of COVID-19 is huge. The health crisis is matched by the worst economic crisis since World War II (Nicola et al., 2020). This dramatic situation is producing a significant effect on personal and social well-being. A study evaluating the psychological impact of the initial outbreak of COVID-19 (Wang et al., 2020) found that 54% of the sample rated the psychological impact of the pandemic as moderate or severe; 17% declared moderate to severe depressive symptoms; 29% reported moderate to severe anxiety symptoms, and 8% reported moderate to severe stress levels. These negative psychological effects are now further aggravated by living in quarantine with its restrictions on movement and social interaction, the difficulties in obtaining basic supplies (e.g., food), and the interruption of professional activities with the subsequent financial loss (Brooks et al., 2020).

How can these problems be tackled? Different countries have opened call centers and online platforms to provide psychological counseling services for patients, their family members, and other people affected by the epidemic (Duan & Zhu, 2020; Imperatori et al., 2020; Liu et al., 2020). However, the organization and management of these psychological interventions have several problems that limit their efficacy (Imperatori et al., 2020). First, patients are not differentiated according to the severity of their situation and often do not find appropriate department or professionals for timely and reasonable diagnosis and treatment. Moreover, the focus of these interventions is for the short term only, with poor followup for treatments and evaluations. Finally, only severe/acute cases obtain psychological interventions, even if the psychological burden of the pandemic has reached the majority of the population.

Could Positive Technology be used for COVID-19? Positive Psychology identifies three characteristics of our personal experience – affective quality, engagement/actualization, and connectedness (Table 1) – that can be manipulated and enhanced to promote personal well-being (Botella et al., 2012; Villani et al., 2016).

To reach this goal, three different types of technologies can be used:

– *Hedonic*: Technologies used to induce positive and pleasant experiences.

– *Eudaimonic*: Technologies used to support individuals in reaching engaging and self-actualizing experiences.

– *Social/Interpersonal*: Technologies used to support and improve social integration and/or connectedness between individuals, groups, and organizations.

Here we explore the potential application of these technologies in addressing the psychological consequences of the COVID-19 pandemic (see the figure below).

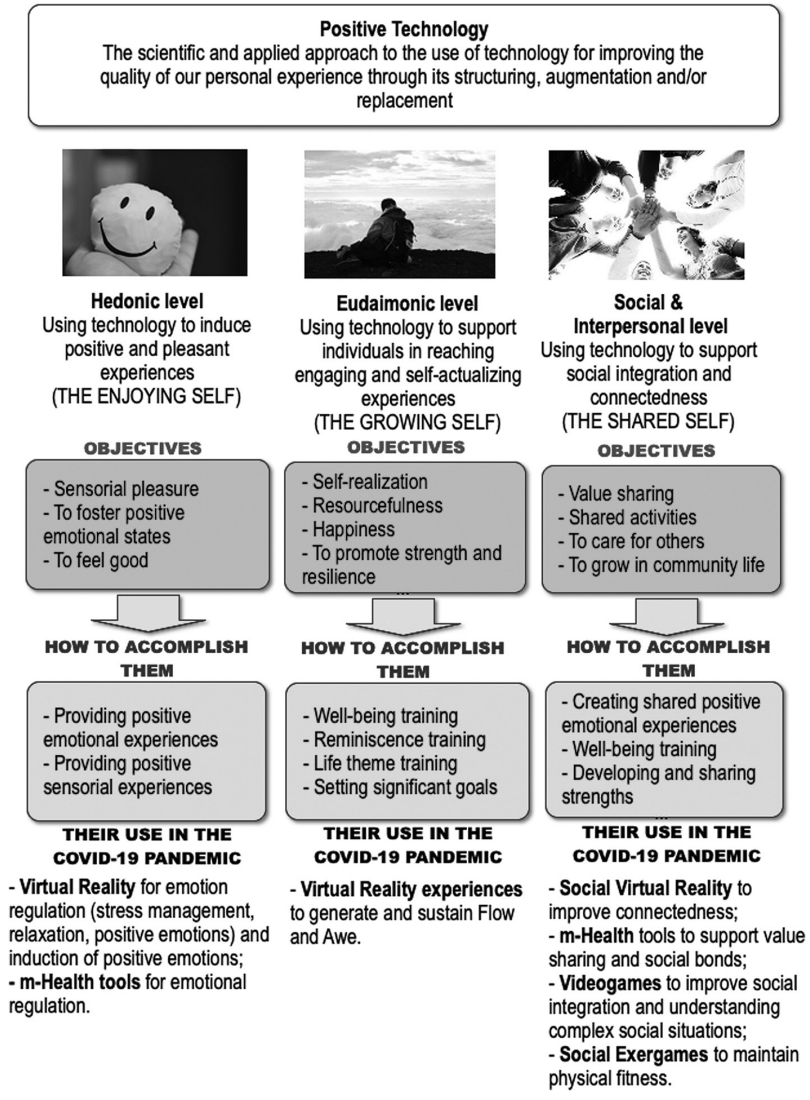


Table 1 - *Personal experience factors for Positive Technology*[*]

Determinants of personal experience	*Key factors*	*Literature and theory*	*Strategies*
Emotional quality (Hedonic level)	Positive emotions	*Building and Broadening Effect* (Fredrickson) *Writing Therapy* (Pennebaker) *Hedonic Psychology* (Kahneman)	Writing therapy Exposure therapy and relaxation Positive ruminating Reframing compassion meditation
	Mindfulness	*Mindfulness-Based Stress Reduction* (Kabat Zinn)	Mindfulness meditation MBSR strategies MBCT strategies
	Resilience	*Psychology of Resilience* (Seligman, Keyes) *Building and Broadening Effect* (Fredrickson)	Positive psychology interventions SuperBetter
Engagement and actualization (Eudaimonic level)	Engagement and presence	*Flow Theory* (Csikszentmihalyi) *Presence* (Riva and Waterworth) *Patient Engagement* (Graffigna, Barello, & Riva)	Challenge and skills Intrinsic and extrinsic rewards
	Self-efficacy and motivation	*Self-Efficacy* (Bandura) *Transtheoretical Model of Change* (Prochaska and DiClemente) *Self-Determination Theory* (Ryan & Deci)	Life summary Online CBT study Technology-mediated reflection

Determinants of personal experience	*Key factors*	*Literature and theory*	*Strategies*
Connectedness (social/ interpersonal level)	Networked flow	*Networked Flow* (Gaggioli & Riva) *Psychological Selection* (Delle Fave, Inghilleri, & Massimini)	Presence and social presence Transformation of Flow
	Gratitude	*Psychology of Gratitude* (Emmons & McCullough)	Gratitude visit Gratitude journal
	Empathy	*Emotional Intelligence* (Salovey & Mayer; Goleman) *Affective and Cognitive empathy* (Gerdes et al.; Singer) *Compassion Focused Therapy* (Paul Gilbert)	Role playing Perspective taking Emotion recognition training
	Altruism	*Empathy Altruism* (Bateson)	Prosocial games Role playing

*Adapted from Villani et al (2016).
CBT, Cognitive Behavioral Therapy; MBCT, Mindfulness-Based Cognitive Therapy; MBSR, Mindfulness-Based Stress Reduction.

Hedonic Technologies: Using Technology to Foster Positive Emotional States

The first dimension of Positive Technology concerns how to use technology to foster positive emotional states. The model of emotions developed by Russell (2003) provides a possible path for reaching this goal: the manipulation of 'core affect', a neurophysiological category corresponding to the combination of valence and arousal levels that endow the subjects with a 'core knowledge' about the emotional features of their experience. Simply put, a positive emotion is achieved by increasing the valence (positive) and arousal (high) of core affect (affect regulation) and by attributing this change to the contents of the proposed experience (object).

Key arguments for the usefulness of positive emotions in increasing well-being have been provided by Fredrickson (2001, 2004) in her 'broaden-and-build model' of positive emotions. According to Fredrickson, positive emotions provide the organism with nonspecific action tendencies that can lead to adaptive behavior (Fredrickson, 2001). For example, in adults, positive emotions make them more likely to interact with others, provide help to others in need, and engage in creative challenges. Moreover, by broadening an individual's awareness and thought – action repertoire, they build upon the resultant learning to create future physical, psychological, and social resources (Fredrickson, 2004).

Several studies have shown that Virtual Reality (VR) represents a highly specialized and effective tool for the induction and the regulation of emotions in both clinical (Carl et al., 2019) and nonclinical subjects (Hadley et al., 2019). In fact, VR can be described as an *advanced imaginal system*: an experiential form of imagery that is as effective as reality at inducing emotional responses (North et al., 1997; Vincelli, 1999; Vincelli, 2001). This feature of VR makes it the perfect tool for stress management applications aimed both to post-traumatic stress and to the generalized stress induced by the pandemic.

An example of this use of VR is the Italian virtual reality for psychological support to health care practitioners involved in the COVID-19 crisis (MIND-VR) project (www.mind-vr.com), aimed at designing, developing, and testing an advanced solution based on the use of VR for the prevention and treatment of stress-related psychopathological symptoms and post-traumatic stress disorder in hospital health care personnel involved in the COVID-19 emergency (Imperatori et al., 2020). In particular, the main goal of MIND-VR is the development of different virtual environments to offer basic education on stress and anxiety and to promote relaxation.

Another possible approach is the use of 'mHealth', the practice of

medicine and public health supported by mobile devices. Although there is no conclusive evidence supporting the efficacy of mHealth interventions, a recent meta-analysis study of 66 randomized controlled trials has confirmed the efficacy of app-supported smartphone interventions on stress, anxiety, depression, and perceived well-being (Linardon et al., 2019).

Different apps – *Virtual Hope Box, Breathe-to-Relax, Calm*, and *Headspace* – developed by reliable sources such as the U.S. Department of Defense and university-based researchers have been suggested to help patients manage anxiety and stress related to the COVID-19 outbreak (Wright & Caudill, 2020). For example, *Virtual Hope Box*, an app developed by the U.S. Department of Defense, includes breathing exercises, deep muscle relaxation, and guided meditation that can help nonclinical populations to address the generalized stress induced by the pandemic.

Eudaimonic Technologies: Using Technology to Promote Engagement and Self-Empowerment

The second level of Positive Technology is strictly related to the eudaimonic concept of well-being, and consists of investigating how technologies can be used to support individuals in reaching engaging and self-actualizing experiences.

The theory of Flow, developed by Positive Psychology pioneer Csikszentmihalyi (1990), provides a useful framework for addressing this challenge. 'Flow' or 'Optimal Experience' is a positive and complex state of consciousness that is present when individuals act with total involvement. The basic feature of this experience is the perceived balance between high environmental opportunities for action (challenges) and adequate personal resources in facing them (skills).

As underlined by Positive Psychology (Delle Fave, 1996; Delle Fave, 2011), to cope with dramatic changes in daily life and to access new environmental opportunities for action, individuals may develop a strategy defined as 'Transformation of Flow' (Riva et al., 2011): the ability of the subject to use Flow for identifying and exploiting new and unexpected resources and sources of involvement. A specific optimal experience that is connected to the Transformation of Flow is the emotional response of Awe (Bai et al., 2017; Guan et al., 2018). In fact, Awe consists of two central features (Bai et al., 2017; Guan et al., 2018) – a) a need for accommodation following the b) perception of vastness – that can induce a significant reorganization of the predictive/simulative mechanisms of the brain:

– *Perception of vastness*: An update of the predictive coding given the

mismatch with current representations of others (Bai et al., 2017; Guan et al., 2018), the world, and ourselves (Newen, 2018), at the base of social order maintenance (Bai et al., 2017; Guan et al., 2018).

– *Need for accommodation*: The tension arising from the mismatch can translate into a drastic update of the predictive coding to accommodate unexpected experiences (Bai et al., 2017; Guan et al., 2018).

A clearer picture of the complex relationship among awe and brain activity was recently provided by a recent voxel-based morphometry study. As demonstrated by Guan et al. (2018) awe involves multiple brain regions associated with cognitive conflict control, attention, conscious self-regulation, and socioemotional regulation, suggesting its potential role in adjusting/improving their functioning.

In this view, awe-inducing technological experiences may be used to induce Transformation of Flow. Among the different types of interactive technologies investigated thus far, VR is considered the most capable of supporting the emergence of both Flow and Awe experiences (Chirico et al., 2017; Gaggioli et al., 2003; Riva et al., 2006; Riva et al., 2010).

The proposed approach is the following (Riva et al., 2018): first, to identify a possible experience that contains functional real-world demands; second, using VR for producing the experience and inducing Awe; third, allowing cultivation, by linking the Awe experience to the actual experience of the subject. The expected effect is a functional reorganization of the brain produced by the broadening of the thought-action repertoire associated with improved self-esteem and self-efficacy.

A practical example of this approach is provided by the free COVID Feel Good (www.covidfeelgood.com) weekly self-help virtual reality protocol (Riva & Wiederhold, 2020; Riva et al., 2020). The protocol consists in watching, for a week, at least once a day, a 10-minute VR video, named 'The Secret Garden', currently available in eight different languages: English, Spanish, French, Brasilian Portuguese, Italian, Catalan, Korean, and Japanese.

Subjects need to watch the video using virtual reality glasses. Then, they need to follow a series of social exercises provided on the project's website, with specific goals for each day of the week.

– Day 1: To prevent individuals from becoming obsessed with coronavirus.

– Day 2: To increase self-esteem.

– Day 3: To work on autobiographical memory (who we are and what we want).

– Day 4: To wake up the sense of community so subjects do not feel alone.

– Day 5: To take back dreams and goals individuals had before the lockdown started.

– Day 6: To work on empathy.
– Day 7: To plan a long-term change.

All the exercises are designed to be experienced with another person (not necessarily physically together), to facilitate a process of critical examination and eventual revision of core assumptions and beliefs related to personal identity, relationships, and goals. Specifically, by facilitating self-reflectiveness and constructive exchange with relevant others, the protocol seeks to improve the ability to adapt to the challenges provided by the pandemic.

Social Technologies: Using Technology to Promote Social Integration and Connectedness

The final level of Positive Technology – the social and interpersonal one – is concerned with the use of technologies to support and improve the connectedness between individuals, groups, and organizations. However, an open challenge is to understand how to use technology to create a mutual sense of awareness, which is essential to the feeling that other participants are there, and to create a strong sense of community at a distance.

As noted by Wiederhold (2020a,b), the coronavirus disease pushed many individuals to shift their work and social lives online. On one side, social media are becoming the most used information tool in disasters such as the current COVID-19 pandemic. On the other side, videoconferencing technologies such as Zoom, Meet, Teams, and WebEx have made it possible to continue some social interaction during quarantine, allowing people to move their lives online while maintaining physical distance to stop the spread of the virus.

However, this process has not been straightforward, and can generate more problems than solutions.

First, an important source of anxiety is the wealth of information that social media provides (Wiederhold, 2020b). According to the World Health Organization (WHO) social media are generating an infodemic, "an overabundance of information – some accurate and some not – that makes it hard for people to find trustworthy sources and reliable guidance when they need it". (Wiederhold, 2020b). Moreover, the increasing use of video calls and meetings is generating a new phenomenon: tiredness, anxiety, or worry resulting from overusing virtual videoconferencing platforms (Wiederhold, 2020a). This technological exhaustion is generated by the technological shortcomings of video calls – delays, lack of eye contact, limited nonverbal cues – that take so much more out of a person than meeting face to face.

How can Positive Technology help in tackling these problems?

One way to overcome technological exhaustion is actually through the use of different technology. Facebook IQ commissioned a study by Neurons, Inc., to compare how 60 participants in the United States responded both cognitively and emotionally – all participants wore EEG headsets to analyze their brain signals and measure their level of comfort and engagement – to conversing in virtual reality versus having a conversation face to face (Facebook IQ, 2017). During the experience individuals met in a virtual conference room appearing as full body avatars: they can fist bump or shake hands and interact with others in ways that make for an experience that is more similar to face-to-face meetings. The results suggest that participants – especially introverts – responded positively to meeting in virtual reality and were able to establish authentic relationships within the virtual environment. As reported by one of the participants of the study (Facebook IQ, 2017): "It was a lot deeper, and enjoyable, and closer to life than I expected. We moved from two strangers having the same (superficial) conversation to two humans revealing themselves and their experience of life". In line with these results several companies have developed and/or recently released different VR social platforms: Facebook Horizon (https://www.oculus.com/facebook-horizon/), VIVE Sync (https://sync.vive.com/login), AltspaceVR (https://altvr.com/), Spatial (https://spatial.io/), and VRChat (https://vrchat.com/).

Multiple researchers have underlined the importance of human values, and the extent to which they are shared by fellow citizens, for tackling the COVID-19 crisis (Wolf et al., 2020). In particular, the awareness that fellow citizens share one's values has been found to elicit a sense of connectedness that may be crucial in promoting collective efforts to contain the pandemic. In this view, technologies that promote online exchanges among individuals across society may also be beneficial for eliciting this sense of connectedness. For example, an international social sharing platform such as 'My Country Talks' (https://www.mycountrytalks.org/) provides a digital place where to set up one-on-one discussions between people with similar or different viewpoints.

Moreover, mHealth offers different tools to improve social connectedness. Banskota et al. (2020) identified 15 smartphone apps that can be used to reduce older adults' isolation. These apps – that range from classical social networking apps such as Facetime and Skype, to apps for visual and hearing impairment such as Be My Eyes – address physical and cognitive limitations and have the potential to improve the quality of life of older adults, especially during social distancing or self-quarantine.

Finally, video games, too, can be used to improve social integration

and connectedness. As underlined by a recent review (Marston & Kowert, 2020), video games allow individuals to connect through play, which is an important source of psychological well-being throughout the lifespan. This feature, combined with the different facets of in-game socialization (i.e., reduced stress, depression, and sense of loneliness) makes video games an important tool for mitigating some of the negative impacts of COVID-19 for adults and children. But the positive potential of video games is broader than this.

First exergames, allowing physical exercise and dance, are important for both maintaining physical fitness and establishing long-term adherence to exercise during the quarantine (Viana & de Lira, 2020). Exergames can also easily be shared with peers and families in social isolation situations becoming a tool for establishing social bonds and collective intentions. Second, video games provide a powerful medium for understanding complex situations, such as the pandemic itself or the fake news associated with it (Kriz, 2020). For example, Factitious (http://factitious.augamestudio.com/) is a simple game that asks players to read a small article and then decide if it is real or fake news (Grace & Hone, 2019). Factitious tries to help players think more critically about fake news by providing contents that make users think critically about the content and the source. The game, by design, aims at the gray area between fake and real news to encourage players to not only think critically, but also practice the work of distinguishing between the two types of content. The new pandemic edition of the game (http://factitious-pandemic.augamestudio.com/) is now specifically targeting the infodemic generated by social media helping individuals to understand what is right and what is wrong. Another interesting example is Plague Inc: Evolved (https://www.ndemiccreations.com/en/25-plague-inc-evolved), which is a real-time strategy simulation video game that allows players to find out more about how viral diseases spread and to understand the complexities of viral outbreaks.

Conclusions

Although the attention of the governments is still focused on the global health emergency produced by the COVID-19 pandemic, individuals are also experiencing extreme psychological stress that is producing a significant burden on our identity and relationships. In this article we discussed the potential of 'Positive Technology', the scientific and applied approach to the use of technology for improving the quality of our personal experience, to augment and enhance the existing strategies for generating psychological well-being. In particular mHealth and

smartphone apps, stand-alone and social virtual reality, video games, exergames, and social technologies have the potential of enhancing different critical features of our personal experience – affective quality, engagement/actualization, and connectedness – that are challenged by the pandemic and its social and economic effects.

In conclusion, as the saying goes, 'a crisis provides an opportunity'; the COVID-19 pandemic provides a great opportunity for promoting and disseminating positive technologies. The *immediate availability* and successful use of positive technologies to tackle a major global societal challenge may serve to increase the public and governmental acceptance of such technologies for other areas of health care and well-being.

References

Bai, Y., Maruskin, L. A., Chen, S., Gordon, A. M., Stellar, J. E., McNeil, G. D., ... & Keltner, D. (2017). Awe, the diminished self, and collective engagement: universals and cultural variations in the small self. *Journal of personality and social psychology, 113*(2), 185.

Banskota, S., Healy, M., & Goldberg, E. M. (2020). 15 smartphone apps for older adults to use while in isolation during the COVID-19 pandemic. *Western Journal of Emergency Medicine, 21*(3), 514.

Botella, C., Riva, G., Gaggioli, A., Wiederhold, B. K., Alcaniz, M., & Baños, R. M. (2012). The present and future of positive technologies. *Cyberpsychology, Behavior, and Social Networking, 15*(2), 78-84.

Brooks, S. K., Webster, R. K., Smith, L. E., Woodland, L., Wessely, S., Greenberg, N., & Rubin, G. J. (2020). The psychological impact of quarantine and how to reduce it: Rapid review of the evidence. *The Lancet, 395*(10227), 912-920.

Carl, E., Stein, A. T., Levihn-Coon, A., Pogue, J. R., Rothbaum, B., Emmelkamp, P., ... & Powers, M. B. (2019). Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials. *Journal of Anxiety Disorders, 61*, 27-36.

Chirico, A., Cipresso, P., Yaden, D. B., Biassoni, F., Riva, G., & Gaggioli, A. (2017). Effectiveness of immersive videos in inducing awe: An experimental study. *Scientific reports, 7*(1), 1-11.

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. HarperCollins.

Delle Fave, A. (1996) Il processo di trasformazione di Flow in un campione di soggetti medullolesi [The process of flow transformation in a sample of subjects with spinal cord injuries]. In F. Massimini, A. Delle Fave, & P. Inghilleri (Eds.), *La selezione psicologica umana* (pp. 615-634). Cooperativa Libraria IULM.

Delle Fave, A., Massimini, F., & Bassi, M. (2011). *Psychological Selection and Optimal Experience Across Cultures: Social Empowerment Through Personal Growth*. Springer.

Duan, L., & Zhu, G. (2020). Psychological interventions for people affected by the COVID-19 epidemic. *The Lancet Psychiatry, 7*(4), 300-302.

Facebook IQ. (2017) *How Virtual Reality Facilitates Social Connection.* https://www.facebook.com/business/news/insights/how-virtual-reality-facilitates-social-connection.

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American psychologist, 56*(3), 218.

Fredrickson, B. L. (2004). The broaden-and-build theory of positive emotions. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 359*(1449), 1367-1377.

Gaggioli, A., Bassi, M., & Delle Fave, A. (2003). Quality of experience in virtual environments. In G. Riva, W. A. Jsselsteijn, & F. Davide (Eds.), *Being There: Concepts, Effects and Measurement of User Presence in Synthetic Environment* (pp. 121-135). Ios Press.

Gaggioli, A., Chirico, A., Triberti, S., & Riva, G. (2016). Transformative interactions: Designing positive technologies to foster self-transcendence and meaning. *Annual Review of Cybertherapy and Telemedicine, 14*, 169-175.

Gaggioli, A., & Riva, G. (2014). Smart tools boost mental-health care. *Nature, 512*(7512), 28-28.

Grace, L., & Hone, B. (2019, May). Factitious: large scale computer game to fight fake news and improve news literacy. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-8). ACM.

Guan, F., Xiang, Y., Chen, O., Wang, W., & Chen, J. (2018). Neural basis of dispositional awe. *Frontiers in behavioral neuroscience, 12*, 209.

Hadley, W., Houck, C., Brown, L. K., Spitalnick, J. S., Ferrer, M., & Barker, D. (2019). Moving beyond role-play: Evaluating the use of virtual reality to teach emotion regulation for the prevention of adolescent risk behavior within a randomized pilot trial. *Journal of Pediatric Psychology, 44*(4), 425-435.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., ... & Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet, 395*(10223), 497-506.

Imperatori, C., Dakanalis, A., Farina, B., Pallavicini, F., Colmegna, F., Mantovani, F., & Clerici, M. (2020). Global storm of stress-related psychopathological symptoms: A brief overview on the usefulness of virtual reality in facing the mental health impact of COVID-19. *Cyberpsychology, Behavior, and Social Networking, 23*(11), 782-788.

Inghilleri P., Riva G., & Riva E. (2015) Introduction: Positive change in global world: Creative individuals and complex societies. In P. Inghilleri, G. Riva, & E. Riva (Eds.), *Enabling positive Change Flow and Complexity in Daily Experience* (pp. 1-5). De Gruyter Open.

Kriz, W. C. (2020). Gaming in the time of COVID-19. *Simulation & Gaming, 51*, 403-410.

Linardon, J., Cuijpers, P., Carlbring, P., Messer, M., & Fuller-Tyszkiewicz, M. (2019).

The efficacy of app-supported smartphone interventions for mental health problems: A meta-analysis of randomized controlled trials. *World Psychiatry, 18(*3), 325-336.

Liu, S., Yang, L., Zhang, C., Xiang, Y. T., Liu, Z., Hu, S., & Zhang, B. (2020). Online mental health services in China during the COVID-19 outbreak. *The Lancet Psychiatry, 7*(4), e17-e18.

Marston, H. R., & Kowert, R. (2020). What role can videogames play in the COVID-19 pandemic?. *Emerald Open Research, 2*, 34.

Newen, A. (2018). The embodied self, the pattern theory of self, and the predictive mind. *Frontiers in psychology, 9*, 2270.

Nicola, M., Alsafi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Iosifidis, C., ... & Agha, R. (2020). The socio-economic implications of the coronavirus pandemic (COVID-19): A review. *International journal of surgery, 78*, 185-193.

North, M. M., North, S. M., & Coble, J. R. (1997). Virtual reality therapy: An effective treatment for psychological disorders. *Studies in Health Technology and Informatics, 44*, 59-70.

Riva, E., Freire, T., & Bassi, M. (2016). The flow experience in clinical settings: Applications in psychotherapy and mental health rehabilitation. In L. Harmat, F. Ørsted Andersen, F. Ullén, J. Wright, & G. Sadlo (Eds.), *Flow Experience* (pp. 309-326). Springer.

Riva, E., Rainisio, N., & Boffi, M. (2015). Introduction: Positive change in global world: Creative individuals and complex societies. In P. Inghilleri, G. Riva, & E. Riva (Eds.), *Enabling Positive Change Flow and Complexity in Daily Experience* (pp. 74-90). De Gruyter Open.

Riva, G. (2012). Personal experience in positive psychology may offer a new focus for a growing discipline. *American Psychologist, 67*(7), 574-575.

Riva, G., Baños, R. M., Botella, C., Wiederhold, B. K., & Gaggioli, A. (2012). Positive technology: Using interactive technologies to promote positive functioning. *Cyberpsychology, Behavior, and Social Networking, 15*(2), 69-77.

Riva, G., Bernardelli, L., Browning, M. H., Castelnuovo, G., Cavedoni, S., Chirico, A., ... & Wiederhold, B. K. (2020). COVID feel good - an easy self-help virtual reality protocol to overcome the psychological burden of coronavirus. *Frontiers in Psychiatry, 11*, 996.

Riva, G., Castelnuovo, G., & Mantovani, F. (2006). Transformation of flow in rehabilitation: The role of advanced communication technologies. *Behavior research methods, 38*(2), 237-244.

Riva, G., Raspelli, S., Algeri, D., Pallavicini, F., Gorini, A., Wiederhold, B. K., & Gaggioli, A. (2010). Interreality in practice: Bridging virtual and real worlds in the treatment of posttraumatic stress disorders. *Cyberpsychology, Behavior, and Social Networking, 13*(1), 55-65.

Riva, G., & Wiederhold, B. K. (2020). How cyberpsychology and virtual reality can help us to overcome the psychological burden of Coronavirus. *Cyberpsychology Behavior, and Social Networking, 5*, 227-229.

Riva, G., Wiederhold, B. K., Chirico, A., Di Lernia, D., Mantovani, F., & Gaggioli, A. (2018). Brain and virtual reality: What do they have in common and how to exploit their potential. *Annual Review of CyberTherapy and Telemedicine, 16*, 3-7.

Riva, G., Wiederhold, B.K., Di Lernia, D., Chirico, A., Riva, E. F. M., Mantovani, F., Cipresso, P., & Gaggioli, A. (2019). Virtual reality meets artificial intelligence: The emergence of advanced digital therapeutics and digital biomarkers. *Annual Review of Cybertherapy and Telemedicine, 17*, 3-7.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review, 110*(1), 145.

Seligman, M. E. (2004). *Authentic happiness: Using the New Positive Psychology to Realize Your Potential for Lasting Fulfillment*. Free Press.

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist, 55*, 5-14.

Viana, R. B., & de Lira, C. A. B. (2020). Exergames as coping strategies for anxiety disorders during the COVID-19 quarantine period. *Games for health journal, 9*(3), 147-149.

Villani, D., Cipresso P., Gagglioli, A., & Riva, G. (Eds.). (2016). *Integrating Technology in Positive Psychology Practice*. IGI Global.

Vincelli, F. (1999). From imagination to virtual reality: The future of clinical psychology. *CyberPsychology and Behavior, 2*(3), 241-248.

Vincelli, F., Molinari, E., & Riva, G. (2001). Virtual reality as clinical tool: Immersion and three-dimensionality in the relationship between patient and therapist. *Studies in Health Technology and Informatics, 81*, 551-553.

Wang, C., Pan, R., Wan, X., Tan, Y., Xu, L., Ho, C. S., & Ho, R. C. (2020). Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. *International journal of environmental research and public health, 17*(5), 1729.

Wiederhold, B. K. (2020a). Connecting through technology during the coronavirus disease 2019 pandemic: Avoiding 'Zoom Fatigue'. *Cyberpsychology, Behavior, and Social Networking, 23*, 437-438

Wiederhold, B. K. (2020b). Using social media to our advantage: Alleviating anxiety during a pandemic. *Cyberpsychology, Behavior, and Social Networking, 23*,197-198.

Wiederhold, B.K., & Riva, G. (2012). Positive technology supports shift to preventive, integrative health. *Cyberpsychology, Behavior and Social Networking, 15*(2), 67-68.

Wolf, L. J., Haddock, G., Manstead, A. S., & Maio, G. R. (2020). The importance of (shared) human values for containing the COVID-19 pandemic. *British Journal of Social Psychology, 59*(3), 618-627.

Wright, J. H., & Caudill, R. (2020). Remote treatment delivery in response to the COVID-19 pandemic. *Psychotherapy and Psychosomatics, 89*(3), 1.

5. Judgements Without Judges

The Algorithm's Rule of Law

G. Della Morte

ABSTRACT

The exceptional aggregate masses of data called 'Big Data', as well as the algorithms that are used to interpret and exploit them, are mostly generated and managed by private companies. Since they are becoming increasingly essential in a wide range of sectors – e.g., the administrative, political, educational ones and, more recently, health and legal environments – this is a matter of concern for the jurists, who are perceiving a transfer of regulatory powers from States and international actors to those in the private sector. In this regard, the main issue that arises, is that the 'algorithm's rule of law' assumes that every problem is computable, and it can be resolved through a finite sequence of well-defined operations. However, law is about principles and values, not numbers, and the issue at stake is precisely how to reconcile the regulatory function of law with the rationale that increasingly underlies the policies based on data computing.

Introduction

Through the Internet "the screen on which people project their lives is no longer and not only that of their personal computer, it has grown considerably and tends to coincide with the entire network space"[1] (Rodotà, 2014, p. 28). This represents a revolution or change at an unprecedented level (Della Morte, 2018; Floridi, 2014; Rodotà, 2021; Smith, 2000; Zuboff, 2018). It was made possible by two main events: the transition from the so-called Web 1 (Free Internet) to Web 2 (Social Web), and that from the latter to the so-called Web 3 (Internet of Things). The first change has allowed the birth of major social networks, the second the development of a communication system through which the same objects – e.g. smartphones or wereables – transmit information useful to obtain, in exchange, services, as in the case of geolocation or contact

[1] Translation of the author. The original text is "Lo schermo, sul quale la persona proietta la sua vita, non è più soltanto quello del personal computer, si è enormemente dilatato, tende a coincidere con l'intero spazio della rete".

tracing apps. The slogan of the German app aimed at counting the development of the COVID-19 epidemic appears, in this sense, paradigmatic: "*Hände waschen, Abstand halten, Daten spenden*" ["wash your hands, keep your distance, donate data"].

The rapid succession of such transformations has resulted in the concentration of aggregate masses of data – the so-called Big Data – "on a scale unthinkable even a decade ago" (Wisnewski, 2016, p. 206). The latter differ in volume, speed, and variety of sources (Mayer-Schönberger & Cukier, 2014), but it is on the qualitative level that the most significant differences are recorded. In fact, the datafication of each experience translates into a radical mutation of our lifestyles, since there is no economic, political, social or cultural sector that is not directly affected. There are those who consider it an attempt to rewrite the world through a new alphabet where letters are replaced by a binary code, an 'ocean of 0 and 1' (Garapon & Lassègue, 2018) that requires special reading tools – algorithms – to be able to be interpreted and generate meaning from an otherwise indistinct mass of information.

Specialists in business, administrative, political, educational and, more recently, health and, especially legal management, are watching in amazement the increasing centrality of data, now as indispensable in public policy planning as they are in defining private strategies (Kittichaisaree, 2017). Whether it be customizing the advertising of a product, developing an app for the health management of pandemic risk, or programming an artificial intelligence capable of calculating the chance of success of a legal argument in court – those reported are all real examples –, algorithms represent irreplaceable allies. The basic legal issue is represented by the fact that although these algorithms increasingly exercise functions of public interest, they are generated and managed by private companies. The topic is of great topical interest and in order to investigate it, Università Cattolica del Sacro Cuore has set up a centre dedicated to these issues (Humane Technology Lab), and a series of University research projects on the impact of algorithms (for the project studying the impact of algorithms in the field of political communication, health and law, see Della Morte, 2020).

A New Type of Raw Material: Data

This abnormal mass of information constantly probed by increasingly complex algorithms is composed of data. They can be classified according to various criteria. For example, they can be personal or anonymous, or even voluntarily entered or recorded autonomously by the network.

The first category (personal data) is defined by article 4 of the Eu-

ropean Union General Data Protection Regulation (GDPR) as: "Any information relating to an identified or identifiable natural person ('data subject')". An identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number or one or more factors specific to, *inter alia*, physical, physiological, cultural or social identity. That means that it does not matter how the data are stored (e.g., on paper or through digital memory). In all cases, personal data are subject to the protection provided by the GDPR. Special categories of personal data are also those that require additional protection because of their ability to identify, *inter alia*, racial or ethnic origin, political opinions, religious or philosophical beliefs, or the health or sexual orientation of an individual.

The opposite of these categories of data are anonymous data which are processed through an anonymization procedure aimed at protecting the privacy of individuals allowing, at the same time, the right development of the social and economic life of a community. It is important to highlight that, under the GDPR provision, personal data that have been de-identified, encrypted or pseudonymized – but can still be used to re-identify a person – remain personal data and fall within the scope of the GDPR. Given the pace of the ongoing technological innovation process, what has just been mentioned is one of the biggest protections offered by the GDPR. It means that for data to be truly anonymized, the anonymization must be irreversible.

A further criterion for classifying data relates to how the data originated. From this perspective, it is possible to isolate four different groups: provided data; observed data; derived data and inferred data (OECD, 2014). The first category (provided data) refers to data that originate from conducts undertaken by individuals who are aware of the actions, e.g. the data shared in social networks – posted data – as well as those produced through payment systems that use credit cards or debit cards – credit data. On the other hand, observed data are not created automatically but include information generated, for example, by cookies or by cameras connected to a facial recognition system. Moreover, both these categories differ from derived data, which are created mechanically by crossing with other data and finally become elements of an aggregate datum referable to an individual. Examples of the latter category include the so-called computational or notational data, used, for example, to understand the relationship between the number of visits to a site and the number of sales of a given product, or to identify the attributes that some groups of buyers have in common for the purpose of classification. Lastly, inferred data represent an analytic-probabilistic process based on a series of correlations, they are currently used for statistical data or advanced analytical data (e.g. credit risk, life expectancy scores).

As can be inferred from this overview, data which we can control directly, those we can decide whether to share or not, are just a small portion of the information that algorithms can process. It is therefore evident that, apart from personal data protection other issues are at stake with reference to derived and inferred data since the people to whom they are addressed are generally not involved in their generation or even not aware of it. It is when the predictive force of the algorithm results into a prescriptive and normative force that bigger issues arise in the eyes of the jurist.

The Disorderly International Normative Framework

The questions that the use of algorithms, that is the systematic exploitation of different categories of data, raises when it comes to rights – from privacy to the protection of personal data, through the compression of fundamental freedoms, such as the expression of thought – are countless. From an overall perspective, the undoubted short-term benefits should be compared to the effects that may settle in a longer-term perspective.

In fact, if on the one hand the forms of governance based on the algorithmic exploitation of data allow more efficient decisions and a safer management of the network, on the other hand they ensure unprecedented forms of control. We had ample evidence of this during the debate around the appropriateness of contact tracing apps. As it has been observed (Harari, 2020) if until yesterday touching your smartphone "the government wanted to know what exactly your finger was clicking on', today 'the focus of interest shifts. Now the government wants to know the temperature of your finger and the blood-pressure under its skin".

Which rules can be invoked to regulate this trade-off between a freer and a safer cyberspace? The problem, far from simple, must be put into the right perspective to avoid slipping into illusory strategies of technosolutionism ("or as Ross Anderson has suggested, 'do-something-it is'", McGregor, 2020). And it must be looked at first of all, as we have suggested in this paper, from the point of view of international regulation, the only one able to face global challenges.

The Internet, or rather the space where data navigate, is a sort of 'network of networks' (the correct expression is Intern-Net-Working), born and flourished in a context completely devoid of public and international control. For this reason, the international regulatory landscape in which to frame the phenomenon of cyberspace is fragmented and articulated, and is more easily understood if two main characteristics are highlighted.

The first concerns the side of subjects or actors. Since there is no in-

ternational organization created to regulate structural aspects, the actors that deal with the Internet at a global level, i.e., above States and regional international organizations such as the European Union, are mostly forums without coercive powers (e.g., the Internet Governance Forum), or they are public subjects *sui generis*, such as the Internet Corporation for Assigned Names and Numbers (ICANN). A non-profit corporation, ICANN, is responsible for the so-called Domain Name System, the search routing mechanism that essentially assigns names to network nodes ensuring that each address is unique and that all Internet users are able to connect to all valid addresses. Originally subject to California law, ICANN has long been closely linked to the U.S. Government's Department of Commerce under a repeatedly renewed Memorandum of Understanding. That status has been the subject of a veritable digital cold war over the years. This led the U.S. Department of Commerce first to recognize it, on September 30, 2009, as a Multistakeholder Governance Model, and then to sign, on March 10, 2016, a new agreement, which came into force the following October 1st and was aimed at ending all relations between the latter and ICANN.

The second relevant feature worth noting concerns norms. Since no general treaty to govern cyberspace has been approved, the international standards that find application are mostly regional, as in the case of the standards developed by the European Union. The General Data Protection Regulation (GDPR), which was approved in 2016 and came into force in 2018, is paradigmatic in this sense. This is a complex body of law developed for the purpose of updating the so-called Privacy Directive adopted more than thirty years earlier and which required a robust update for two different reasons. First of all, because in the last three decades technological innovation has affected our lives more and more profoundly, in a previously unimaginable way; and, secondly, because in the same period legal sensibilities have changed. It is in this perspective that it is necessary to interpret the distinction, introduced by the EU Charter of Fundamental Rights, between the right to 'privacy' and 'protection of personal data'. The former, conceived to cope with the spread of photographic technology (Brandeis & Warren, 1890), is defined as the right to be let alone because it is inspired by the idea that we own all the information concerning ourselves. On the other hand, the need to protect personal data emerges in a world where the traffic of information has become an unavoidable factor of development: the aim is therefore to allow its circulation, while ensuring control over personal data and especially sensitive personal data that require special protection (such as those that detect racial or ethnic origin, religious beliefs, etc.).

In addition to the latter provisions, a number of other rules created at different times and in different circumstances may also apply. Exam-

ples include sector-specific conventions that directly or indirectly protect privacy and personal data from specific perspectives, such as the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data adopted in 1981 and that was the first legally binding international instrument in the data protection field; or treaties that were meant to regulate different issues like the United Nations Convention on the Law of the Sea adopted in 1982 and regulating, *inter alia*, the right to lay submarine cables to carry our data.

In such a context, it is difficult to imagine permanent, *prêt-à-porter* solutions. All the more so since the regulatory landscape is in continuous movement, as witnessed by the proposal for a Regulation of the European Parliament and the Council laying down harmonised rules on artificial intelligence (2021).

Moreover, on a global scale, we are witnessing a clash between three major players, the European Union, the United States of America and China, which are characterized by different sensibilities on the issue of personal data protection.

Those differences, which are very evident with China, are recognizable even just between the two sides of the Atlantic Ocean. On the one hand, Europe, more attentive to the protection of rights; and on the other, the USA, more sensitive to the exercise of business freedom. The simple tension that had originally opposed the positions of these two players in relation, for example, to the regulation of Passenger Name Records for air transport, exploded following the Datagate case first, and the judgements of the Court of Justice of the European Union in the case Google *v.* Spain (2014) and in the case Schrems *v.* Facebook (2015).

Even the so called 'Privacy Shield' – an international agreement on transatlantic data flows signed by the European Union and the United States in 2016 in order to replace the former 'Safe Harbour' – has subsequently been declared invalid by the Court of Justice of the European Union. In a new judgement adopted on July 2020 the same Court considered that the requirements of U.S. domestic law result in limitations on the protection of personal data equivalent to those required under EU law, and that this legislation does not grant data subjects actionable rights before the courts against the U.S. authorities (Schrems *v.* Facebook, Part II).

The Main Legal Problem Concerning the Algorithm's Rule of Law

That having been noted, the main problem concerning content regulation in cyberspace is that data mining is increasingly clashing with the principles set out in international human rights law. The prediction of

similar 'tension axes' relies upon the idea that the logic supporting the entire sequence of creation, preservation, transfer and possible elimination of the various contents (data) available online is inspired by a general assumption: every computable problem can be solved through an algorithm. This leads to a general issue: while, on the one hand, fundamental human rights normally favour disadvantaged individuals as they naturally support minorities, on the other hand, big data is more inclined to the largest number, strengthening the majority, thus increasing prejudice against the minority (the so-called Dictatorship of Data). But "law is about values, not about numbers" (Zeno-Zencovich, 2017, p. 13) and the problem is precisely how to reconcile the regulatory function of law – often expressed in forms of minority protection – with the rationale that increasingly underlies the policies based on data computing – which conversely expresses the direction of the majorities. It is the case of a software like the so-called COMPAS – Correctional Offender Management Profiling for Alternative Sanctions – an algorithm used by U.S. courts to assess the potential recidivism risk through a scale based on prior arrest history, residential stability etc. In July 2016, the Wisconsin Supreme Court ruled that COMPAS can be only considered, along with "other independent factors", by judges during sentencing, but cannot be 'determinative' (State of Wisconsin *v.* E. L. Loomis, 2015AP157-CR, 13 July 2016, para. 9). This kind of example raises the question of how to solve the contrast between the predictive function of the algorithm and the prescriptive function of law. The contrast is more than theoretical and lies in the capacity of the law to make a decision and to motivate it. It also lies in the inability of the algorithm to decide and elaborate rational arguments. In sum the algorithm does not interpret but limits itself to processing data. If the main concern is to preserve the nature of human rights – all based on an assumption of dignity: humankind as an end to itself and not in relation to other elements – then the principle of causality upon which the legal reasoning is built must be kept clearly distinct from the principle of correlation on which algorithms are shaped.

The issue is made even more sensitive by the fact that we cannot exclude the application of big data in favor of human rights protection, for instance giving a more precise indication in medical decision-making. By improving prediction, the use of artificial intelligence can optimize resources allocation, support socially and environmentally beneficial outcomes and provide advantages in key sectors like environment, health, agriculture, mobility, and home affairs. Still, the same techniques can also bring new risks or negative consequences for individuals or the society. Therefore, the question arises whether the law provides mechanisms for resolving similar conflicts.

In other words, the issue is no longer whether or not a possible algorithmic system is capable of administering power effectively: it certainly can. In fact, the problem of governance without government of algorithms lies in the absence of that system of checks and balances stratified over the centuries, as well as in the circumstance that the control of the balance between rights equally deserving of protection is managed by private subjects. Consider, lastly, the Facebook Oversight Board: an unprecedented experiment in private justice hypothetically applicable to billions of users. In it, the private exercise of public functions appears barely mitigated by reference to norms and principles of international protection of human rights (Gradoni, 2021; Klonick, 2020; Pollicino, 2021). In an optimistic perspective, it is only a problem of evolutionary adjustment; in a pessimistic one, it is the antechamber to new, nefarious, uncontrolled concentrations of power.

Judgements Without Judges?

Given these considerations, it seems undeniable that the entry of digital governance brings about a political, sociological and cognitive mutation. Today, individuals are more likely to believe the results of an algorithmic function than an interpretation. Although such an observation is causing more than one 'narcissistic wound' among old school professionals, it deserves to be addressed with the greatest seriousness. On the other hand, it is hard to imagine that we can go backwards. In fact, after healthcare, whose digital transition has undergone an extraordinary acceleration due to the COVID-19 pandemic, the law could be the next territory colonized by digital technology.

What might happen if one day judges were replaced by simple apps that make use of predictive algorithms? If contracts were signed, challenged and resolved on online platforms? If transcripts in public records were supplanted by blockchain recording systems? In short: how smart is artificial intelligence? Enough to be able to be used in that delicate experience that goes by the name of trial? And, if so, who can determine that? A judge, or, once again, an algorithm?

It is tempting to respond negatively. Yet, if an ordered sequence of computational operations is faster, more powerful and unbiased than human reasoning, one can infer that the algorithm is better suited to process complex decisions such as those expressed in the process (we are referring here to so-called replacement automation – that is machines replacing humans – and not simply to the support function, on which there is general agreement; Nieva-Fenoll, 2018). Nor would the objection be valid that the algorithm does not feel, and therefore, not

knowing empathy, cannot make a correct judgement. The algorithm can learn empathy based on scales elaborated on measurable values; it can, in other words, *pretend* to feel. In fact, artificial intelligence proceeds by *mimesis*, by copying human behaviour. So it would be enough to elaborate measurement criteria for each of the elements to be weighed – fear, emotion, anger, repentance, etc. – to overcome this first objection. After all, if a simple smartphone recognizes the shortest or least busy route to our destination, why could not legal software identify the preferable argumentative path or the most significant precedents to allow us to make the best decision? We are not far from such scenarios. Alibi software, for example, illustrates the arguments that someone charged with a specific crime can make; Watson Debater recognizes the most commonly used arguments with respect to a topic of discussion; Ross Intelligence suggests the correlation between legal reasoning and success rates in litigation, and so on.

There remains, however, the knot of judgement. When we consider how algorithms can help us to make better judgements, we forget to mention better for whom, or in relation to what. It is necessary to focus every attention on this point, also because we are at the dawn of a technological revolution that – just because it is revolutionary – has the character of irreversibility.

In this regard, it should be stressed once again that the *modus operandi* of artificial intelligence is based on the use of algorithms, complex sequences of calculation able to extract from the indistinct mass of big data the information or correlations relevant to solve the concrete problem we are dealing with, whether it be the quantification of the maintenance allowance in case of divorce or the calculation of the danger of recidivism. Since the object of this type of operation is the treatment of data and not an interpretation process starting from general principles, the result is an inevitable flattening of the legal activity here reduced to a mere calculating procedure. As mentioned, such a downgrade is not without risks. But these risks do not manifest on the level of fairness – it is certainly easy to imagine an algorithm fairer than a judge – but on that, much more general, of the false assumption for which past and present are commensurable experiences, weighable on the same level.

The jurist, who like the historian must reconstruct the past on the basis of simple traces, knows well that men ‘make history’, but ‘do not know the history they make’. This is because history also feeds on discontinuity, and sometimes it is precisely the interpretations of rupture that trigger spirals aimed at strengthening interests until then unprotected. The algorithm is not able to understand all this, if not recording first, and copying then, a sequence of discontinuity that in this way would become in turn predictable and continuous. In such a scenario,

the hemicycle-shaped classroom where everyone can see and everyone can simultaneously hear (because "a word is also understood through its own silences, as well as through the silences it is capable of producing: 'a silence of death'"; Garapon & Lassègue, 2018, p. 185) appears outdated. In the digital justice model, the three units of time, place and action – adopted in a traditional trial and inspired by classical theatre – are reduced to one, because there is no need for a temporal succession in which the parties alternate in the presentation of arguments, nor for an overall perspective.

The algorithm decides on the basis of correlations, it does not need principles from which to make the law descend, through reasoning. But these correlations are based on past events: the algorithm, in other words, is unsurpassed in 'predicting the past'. However, on the other hand, there are areas where law is naturally unbalanced in favour of the weak. This is particularly evident in the context of the protection of human rights, where the use of decisions based on a predictive perspective (ergo on the basis of big data) can, as evoked above, easily come into collision with the prescriptive nature of a law fundamentally born to protect minorities against the abuses of majorities.

I use the word 'still' because it is clear to me that we are at the dawn of a new paradigm, in respect of which the warning expressed by Roscoe Pound in 1923 remains valid: "Law must be stable and yet it cannot stand still" (p. 1). Certainty of the norm *versus* adaptability. The playwright of law is still all there.

Conclusion

The first law of geography has been described in the following terms: "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970, p. 236). But since the computer of the University of California Los Angeles (UCLA) connected with that of the Stanford Research Institute, in 1969, this law was already 'bound to be obsolete' (Vegetti, 2017). In fact, cyberspace implies a revolution in the concept of proximity and distance. This revolution acts in the first place at the spatial level, but it concerns also the temporal dimension, as we have just recalled. Considering that the Internet protocol already connects, as of March 31, 2021, 5 billion and 168,000 Internet users all around the world (Internet World Stats, 2021), it is evident to what extent the transition from the analogue to the digital world is bound to have an impact on our lives. After all, Alphabet (Google), Amazon, Apple, Facebook and Microsoft are not only Big Tech companies with mega-capitalization in a sector so strategic as to have obscured the role of

Big Oil in the mining sector (with the difference that it is data that have replaced oil). They are also asserting themselves as centres of power exercising public functions. Presiding unchallenged over entire areas of cyberspace, every time they modify their computer code, they produce norms. Think of the rules on access to digital data in the event of death: establishing what our digital identities will be able to do after our passing means establishing a kind of inheritance law, potentially applicable to billions of individuals.

As we have observed, such a profound transformation is not without its criticalities. First of all, algorithms are not yet able to interpret, i.e., to think in the common sense of the term, but are limited to correlating data. Moreover, although they exercise functions of increasingly public interest, algorithms are often produced by companies listed on the stock exchange. It is hardly worth remembering how, in the first days of 2021, the former President of the United States of America was essentially silenced by the major social networks, after stirring up the most radical wing of his supporters with some tweets. During the unfortunate attack on the Capitol in Washington, Trump's profile was first blocked by Twitter and Facebook and then by Twitch, Reddit, Shopify, Snapchat and, to a lesser extent, by TikTok and Pinterest. The episode caused a stir, and sparked a passionate debate: can access to the public digital arena, the only marketplace that can be frequented in these times of global pandemic, be subjected to private control? How can we be sure that they pursue a collective interest? *Quis custodiet ipsos custodes*? If the answer does not consist in restricting the field of the free manifestation of thought, but in enlarging the enclosure of the free market of ideas, the solution does not appear easy to find, precisely because of the concentration of power mentioned above.

Moreover, when, after the block, Trump tried to move to Parler, a social platform popular in right–wing radical circles, Apple and Google immediately excluded the possibility of downloading the relative app and Amazon suspended the social network from its servers. It's not just access to the public digital arena that is privately managed. It's also the infrastructure on which the network relies.

How do jurists, and in particular internationalists, deal with such changes? As recalled, since there is no general treaty on the rights and duties of cyberspace, nor an international organization invested with such powers, the attention is mostly focused on norms elaborated at regional and sectoral level, with particular reference to the theme of personal data protection elaborated within the European Union and refined by the jurisprudence of domestic and international courts.

Against the backdrop of this consideration, it is also clear that the sensitivity shown by the European Union on these issues is due as much

to economic reasons as to historical-social reasons. On the first level, it should be remembered that European consumers' data are particularly attractive on the global market: referring to a population of almost four hundred and fifty million individuals, with life expectancy and consumption power well above the global average, it is understandable that there is interest in conserving data within the common market. Secondly, we must not forget that European institutions are based on the horror of Auschwitz, the Holocaust made possible thanks to a distorted use of personal data. Just think of the triangles of various colours used by the Nazi regime to classify individuals in the concentration camps. The classification criteria that referred to racism, religion, sexual orientation, etc. are what we now call special categories of personal data under article 9 of the GDPR.

It is also in this light that the efforts of the European legislator must be appreciated: the discipline of personal data protection is rooted in the deepest foundation of the European Union.

However, how well the effort rises to the challenge remains an open question. Some positive signs come from privacy by design and by default: an innovative approach that requires companies to start a project or create a device by providing, from the outset, the correct settings to protect personal data. But in other ways the answer is negative: even the GDPR, although very recent, is partly outdated. Especially when it takes the form of practices such as requests for authorization, etc. that do not keep pace with technological innovation and with the circumstance that, as explained above, the data that matter most today are those derived or inferred, and not the provisional data that we voluntarily introduce into cyberspace.

In conclusion, the identification of a correct set of principles would mean building a compass capable of guiding us in the difficult navigation of the deep waters of cyberspace. This image is not a coincidence: the etymology of the word -cyber itself comes from the ancient Greek (*kyber*) meaning 'helm'. Thus, the problem is twofold: how to steer and who is at the helm.

'How to steer' is to be determined by certain principles of law, consolidated and being currently defined by domestic and international courts. However, at the moment private companies are likely to be 'at the helm', as witnessed by the case of the legal dispute generated between the US Government and Apple following the San Bernardino massacre. This is an emblematic case and for this reason we feel it is appropriate to recall it. On December 2, 2015, the FBI requested Apple to access the contents of an iPhone found in the possession of a bomber who had died in the terrorist attack in San Bernardino, California. The reason was simple: the terrorist's iPhone – Model 5 – was the first

in the series not to be equipped with a backdoor, or software to circumvent the data protection provided by the device through privacy by design. However, the information contained (messages, traces, etc.) was related to national security issues and the FBI believed it had the right to access it. The result was a dispute between FBI and Apple about the encryption of the device, which had a classic problem of legal balance (between security and protection of personal data) as an object and which ended with the refusal of the Cupertino company. The denial letter deserves to be mentioned because it could serve as an introduction for a future course on the transformations of sovereignty: according to Apple, the US government had requested something that 'in the wrong hands' would have determined an excessive concentration of power, since it would not have been possible to verify if the authorities had made a use limited to the case.

An excessive concentration of power that in the wrong hands would have been too great a danger. It's a pity that in the context the meaning appears reversed in favor of a private company: it could not have been said better.

References

Brandeis, L., & Warren, S. (1890). The right to privacy. *Harvard law review, 4*(5), 193-220.

Della Morte, G. (2018). *Big Data e protezione internazionale dei diritti umani. Regole e conflitti* [Big Data and international human rights protection. Rules and conflicts]. Editoriale scientifica.

Della Morte, G. (Principal Investigator) (2020-2023), Funzioni pubbliche / controllo privato [Grant]. Università Cattolica del Sacro Cuore. https://progetti.unicatt.it/fpcp-home-presentazione.

Floridi, L. (2014). *The Fourth Revolution. How the Infosphere is Reshaping Human Reality.* Oxford University Press.

Garapon, A., & Lassègue, J. (2018), *Justice digitale. Révolution graphique et rupture anthropologique.* Presse Universitaire de France.

Gradoni, L. (2021). Constitutional review via Facebook's Oversight Board: How platform governance had its Marbury v Madison. *Verfassungsblog (blog).* 10 February 2021. https://verfassungsblog.de/fob-marbury-v-madison/.

Harari, Y. N. (2020). The World after Coronavirus. *The Financial Times*, 20 March. https://www.ft.com/content/19d90308-6858-11ea-a3c9-1fe6fedcca75.

Internet World Stats (2021). Internet in Europe Stats. www.internetworldstats.com.

Judgment of the Court (Grand Chamber), 13 May 2014, Google Spain SL and

Google Inc. v. Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, ECLI identifier: ECLI:EU:C:2014:317. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:62012CJ0131&from=EN.

Judgment of the Court (Grand Chamber) of 6 October 2015, Maximillian Schrems v Data Protection Commissioner, ECLI identifier: ECLI:EU:C:2015:650. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:62014CJ0362&from=en.

Judgment of the Court (Grand Chamber) of 16 July 2020, Data Protection Commissioner v Facebook Ireland Limited and Maximillian Schrems, ECLI identifier: ECLI:EU:C:2020:559. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:62018CJ0311&from=en.

Kittichaisaree, K. (2017). *Public International Law of Cyberspace*. Springer.

Klonick, K. (2019). The Facebook Oversight Board: Creating an independent institution to adjudicate online free expression. *The Yale Law Journal, 129*, 1418-1499.

Mayer-Schönberger, V., & Cukier, K. (2014). *Big Data. A Revolution that Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt.

McGregor, L., (2020). Contact-tracing Apps and Human Rights. *EJIL-Talk! – Blog of the European Journal of International Law* (www.ejiltalk.org), 30 April. https://www.ejiltalk.org/contact-tracing-apps-and-human-rights/.

Nieva-Fenoll, J. (2018). *Inteligencia artificial y proceso judicial* [Artificial intelligence and judicial process]. Marcial Pons.

OECD (2014). Working Party On Security And Privacy (doc. DSTI/ICCP/REG(2014)3). https://www.oecd.org/digital/ieconomy/workingpartyonsecurityandprivacyinthedigitaleconomyspde.htm.

Pollicino, O., & De Gregorio, G. (2021). The first decisions of the Facebook Oversight Board. *Verfassungsblog (blog)*. 5 February 2021.

Pound, R. (1923). *Interpretations of Legal History*. MacMillan.

Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain union legislative acts, Brussels, 21.4.2021 COM(2021) 206 final. https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735–a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF.

Rodotà, S. (2014). *Il mondo nella rete – Quali diritti, quali vincoli* [The world in the net – Which rights, which constraints]. Laterza.

Rodotà, S. (2021). *Tecnologie e diritti (2° Ed.)* [Technology and rights (2nd Ed.)]. il Mulino.

Smith, L. (2000). *The Third Industrial Revolution: Law and Policy for the Internet*. The Hague, Collected Courses of the Hague Academy of International Law.

State of Wisconsin v. E. L. Loomis (881 N.W.2d 749, Wis. 2016)), para. 9. https://www.wicourts.gov/sc/opinion/DisplayDocument.pdf?content=pdf&seqNo=171690.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography, 46*(sup1), 234-240.

Vegetti, M. (1927). *L'invenzione del globo. Spazio, potere, comunicazione nell'epoca dell'aria.* Einaudi.

Wisnewski, J. (2016). WikiLeaks and whistleblowing: Privacy and consent in an age of digital surveillance. In J. Galliot, & W. Reed, *Ethics and the Future of Spying: Technology, National Security and Intelligence Collection* (pp. 221-232). Routledge.

World Economic Forum Report (2011). *Personal Data: The Emergence of a New Asset Class.* https://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf.

Zeno-Zencovich, V. (2019). *Ten Legal Perspectives on the "Big Data Revolution".* Editoriale Scientifica.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future and the New Frontier of Power.* Public Affairs.

APPENDIX I

Robotization of Life: Ethics in View of New Challenges

ABSTRACT

The document *Robotization of Life* has been produced by an *ad hoc* working group on robotization established by the Commission of the Bishops' Conferences of the European Union (COMECE). Led by Professor Antonio Autiero and enriched by diverse contributions of experts in theology, philosophy, law and engineering, the COMECE working group analyzed the impacts of robotization on the human person and on society as a whole and elaborated its reflection as an ethical step which can shape community life in our complex and globalized society in which actors are increasingly interconnected. The document reaffirms the primacy of the human, on the basis of the recognition of the human dignity of each person.

Introductory Remarks

Scope of the analysis

The development of robotization is linked to a number of factors.

1. Given the complexity of tasks to be performed in a society, there is increasing reliance on sophisticated technological tools (for communication, transport, information processing, etc.). These exceed the speed and precision of human actions and reactions, as well as memory and perception capabilities. In a complex and globalized society of increasingly interconnected actors, robotization transcends human physical and cognitive limits in decision-making and regulation processes.

2. Robotization furthers the aim of minimizing production and labor costs.

3. Robotization reduces dangers to which workers are exposed. This is particularly the case in potentially dangerous industries, as well as policing and the military[1]. The desire to increase performance and the

[1] *The Humanization of Robots and the Robotization of the Human Person. Ethical Reflections on Lethal Autonomous Weapons Systems and Augmented Soldiers* (with a selection of texts from the

profitability of processes in an increasingly complex and technology-enhanced society has led to (at least in societies that can afford sophisticated technological means) the gradual replacement of the human person by the machine.

Robotization has already the capacity to significantly help the medicine sector by recognizing and detecting diseases through fast, efficient and standardized means. It also makes possible to offset handicaps (for example, exoskeletons, prostheses, etc.), to administer treatments automatically and to carry out surgery with a high degree of accuracy and precision, and to do this remotely. Despite the advantages of robotization, it should be noted that it has developed within a culture that no longer tolerates the limits of the human person. Projects involving robot-assisted persons, or robotized (or augmented) human persons, are motivated by the desire to free humanity from biological constraints (for example, physical resistance, mental capacities, ageing, etc.) so as to be master of its being and becoming. Admittedly, this falls short of the 'trans-' or 'post-humanist' utopian philosophies which permeate certain spheres of contemporary thought. Robotization is nevertheless associated with and motivated by the idea that the human person is able to transform itself so as to escape its limited, fragile, biological condition, a condition which is considered unbearable and therefore to be overcome.

4. Robotization develops in the context of the 'anthropological crisis', understood to be a radical questioning of the identity and true reality of the human person. The intensification of robotization, with the consequent redundancy or transformation of the human person, has implications for particular societies and certain population groups. Some societies cannot afford efficient robotization and certain populations groups or classes which, because of economic reasons or personal physical or mental disabilities, are left behind due to their lack of access to technologies. The robotization of life must therefore be wisely and critically considered as an opportunity but not as an absolute necessity (because it is related to certain particular interests) and with a concern for those potentially left behind. It must also be noted that robotization is, in certain sectors, driven by factors which are themselves reinforced by the robotization they have created[2].

Holy See's engagement on Lethal Autonomous Weapon Systems), Geneva, The Caritas in Veritate Foundation Working Papers, vol. IX, 2017.

[2] Lambert, D. (2013). Risques et espoirs d'un discours sur la vulnérabilité humaine. In *Fragilité, dis-nous ta grandeur.* Cerf (pp. 13-30). Proceedings of the conférence: *Sense or nonsense of "Human Fragility" in contemporary European society,* European Parliament, Friday 21st of October 2011; Id., Faut-il se libérer de la fragilité? Questions posées par une robotisation des activités humaines. In *Fragilité, dis-nous ta grandeur.* Cerf (pp. 101-118).

Terminological clarification

Digital data systems enable intelligent processes to be replicated to an even greater degree. The term, 'artificial intelligence' is used as a generic term for these systems. This paper's focus is the specific processes carried out by robots. A robot is a system that usually consists of three components: 1) a sensor which gathers information from its surroundings; 2) a processor, which processes that information; and 3) an effector which can interact with the surrounding environment.

Diversity and specificity of the ethical issues

The ethical considerations that arise in the context of robotization are relevant more generally to the relationship between science and ethics. It must be acknowledged that the development of technology provides necessary support to individuals and society in the exercise of human responsibility. To that end, technological advances must not be demonized or rejected. What is necessary is a focused ethical analysis of the impact of accelerated and advanced process of robotization on the individual and society.

Ethical Issues

Primacy of the person, recognition of the human dignity

Some contemporary scientists and philosophers claim that robots have a certain degree of autonomy in the sense that they are subjects that act. They can therefore be considered within certain limits as so called 'moral agents' in that they can make choices that can be assessed as good or evil. This would give rise to ethical problems.

During life, a human person can find themselves interacting with these robotic agents. No longer is the human-robot relationship defined in instrumental terms, with use of the robot strengthening human action. Instead, if human persons seek to exercise control over their environment, they need to confer power on other entities, in this case, artificial entities. This requires accepting that, because of the increased autonomy and agency of such entities, human action is limited. At the same time, actions over which humans have control increase. This gives rise to a paradox: the more human power over the environment increases thanks to machines, the more human beings are deprived of agency and control.

This paradox generates a sense of unease and powerlessness. The dignity and centrality of the human person is put into question. It is therefore necessary to extend the principle of good relations, which previously regulated human interaction with nature and other human beings, to include robots.

In this regard, two steps must be taken, both of which are grounded in the idea of '*creaturality*'.

First, just as human persons, in their freedom, deliberative decision-making process and autonomy, are creatures of God; so robots, despite their 'autonomy', are designed and programmed by humans. A human person and a cognitive machine have specific capacities to start processes; they can relate to and interact with each other; and most importantly, their activity can be subject to moral judgment and assessment as to whether it is good or evil – in the sense of harmful or harmless activity. Nevertheless, the machine only acts according to its original programming by a human person. Then, even if the machine can interact with and even assist human persons, it is not properly a moral agent and the ultimate responsibility always lies on humans.

Second, and most important, what governs the relationship between humans and machines is the primacy and dignity of the human person. Although created, the human person is not only capable of relating on his or her own to other creatures (just as robots are, to a certain degree, also programmed to do), but also has the capacity to question the criteria and principles upon which to make decisions. The human person is capable of critical reflection and ethical decision-making, like Adam in the Garden of Eden (see *Gn* 2).

The human person is responsible for giving order and meaning to creation. Christian anthropology, itself rooted in the wisdom of the Judeo-Christian biblical tradition, articulates and develops a vision of the human person whose primary task it is to preserve and cultivate nature. This grounds an ethics which does not idealise nature in a sacral or romantic sense. It goes beyond mere preservation to practically cultivate, develop and increase creation. This dynamic sense of humanity's role in creation supports not a conservative ethics, but rather a future-oriented one which is open to and responsible for creation as it grows and develops. This promotes an attitude toward science and technology which is fundamentally confident and welcoming of innovation.

Moreover, it emphasizes the value a person's freedom and non-dependence upon the technology at their disposal. This is expressed in terms of a person's critically reflective and evaluative attitude to the use (or misuse) of technology.

The robot, at least in its present phase of development, is not capa-

ble of this. It can only follow the procedures for which it has been programmed. As a consequence, only the human person can be considered a 'person' in the proper sense, and in their full dignity.

Rights of robots

The wide and varied range of ethical challenges which arise from society's use of robots comes to the fore in the ongoing debate as to whether robots should be accorded specific legal status and attendant rights.

The European Parliament has recommended this in its *Resolution on Civil Law Rules on Robotics*[3]. It proposes that the most sophisticated, autonomous robots be given the status of 'electronic persons', responsible for making good any damage they cause. It further recommends that '*electronic personality*' be applied to cases where robots make autonomous decisions or otherwise interact with third parties independently.

It must be said, however, that the construct of legal status for robots is unconvincing. The human person is the foundation of, and central to every legal order. For a natural person, legal personality derives from their existence as human. That personality implies rights and duties that are exercised within a framework which recognizes, respects and promotes human dignity. Placing robots on the same level as human persons is therefore at odds with Article 6 of the *Universal Declaration on Human Rights*, which states that "*everyone has the right to recognition everywhere as a person before the law*".

Calls for the extension of legal personality to robots runs contrary to, and undermines the very concept of responsibility, as it arises in the context of human rights and duties. Responsibility rooted in legal personality can only be exercised where there exists the capacity for freedom, and freedom is more than autonomy.

Legal personality is assigned to a natural person (as the natural consequence of their being human) or to a legal person (in this case, even though a fiction, legal personality presupposes the existence of a natural person or persons acting behind the fiction). Legal personality for robots collapses the boundaries between humans and machines, between the living and the inert, the human and the inhuman[4].

Some contend that rules of liability could be extended to robots in a way analogous to the rules which govern liability associated with ani-

---

[3] European Parliament (2017). *Resolution on Civil Law Rules on Robotic.*

[4] *Contribution of COMECE to the public consultation of the European Parliament Civil Law Rules on Robotic*, 2017.

mals. This would represent a perilous shift towards the recognition of robots as belonging to the world of the living. It remains that existing legal frameworks which provide for natural and legal personality already have at their disposal viable legal solutions, not least provisions on defective products, as well as rules about the liability for damages or injury caused by things in a person's care.

Particular Focus

How the future of work will change?

The fields of application for robotics are many and varied. Some ethical problems emerge in relation to particular areas of application, whilst others are basic and remain common to all. As has been noted, a field which calls for particular attention is undoubtedly the labor market and the personal and societal impact of robotization. The development of the labor market, and the prospect of increased human redundancy makes it a controversial topic for consideration.

The use of robots will cause profound societal change. This will be most obvious in the context of the labor market where conditions are likely to undergo radical change. Robots will be able to extend, even replace work previously done by human persons. This phenomenon has been described as the Fourth Industrial Revolution and it is ongoing, shaping significantly current and future employment patterns[5].

Studies also predict enormous change in job profiles. In order to integrate robots, the work environment requires reorganization and restructuring which itself generates new jobs which differ from existing employment profiles. An advantage of the use of robots in these new jobs is that it minimizes human exposure to dangerous and inhumane working processes.

It must also be noted, however, that whilst robots in the workplace bring with them opportunities and advantages, they also (often adversely) affect the most vulnerable groups in society, particularly young people and the less well educated.

Robots can easily perform simple, automated sequences of work which traditionally were carried out by young workers entering the labor market, or by the unskilled. This has the potential to lead to a decrease in job security for these groups and increased polarization of the labor market.

The needs of contemporary society call for a renewed commitment

[5] COMECE (2018). *Shaping the future of work.*

to the shaping and regulation of the use of robots in the workplace. This requires that legislators be attentive to a number of factors: the safety of the labor market has to be assured, the common good must be respected and the rights of workers need to be protected.

The existing European legal framework states that work is a human right and that favorable working conditions must be provided. Human dignity, individual freedom and solidarity are foundational to these rights, and they give rise to the obligation to shape a human-centered view of work for the future.

How social justice and common good become decisive ethical criteria?

Any ethical analysis must be conducted cognizant of both – the individual and collective perspectives. The moral and ethical responsibility to be exercised in the use of robotics relates not only to the primacy of the individual, respect for their dignity and the safeguarding of their free choices, but also wider considerations of social justice.

Social justice is not solely concerned with the end-goal of the common good, but with issues of equitable distribution and fair access to the world's resources, and here, robotics has a role. The danger with the growth and development of robotics is that already-existing social differences are being exacerbated, injustices and inequalities are increasing (especially for the most vulnerable) and the attainment of the common good is being frustrated.

The Christian anthropological vision is one based on solidarity and itself provides a basis for minimizing, even overcoming, the negative impacts of robotics, especially for the poor. The idea of the common good is not an abstract one. Rather, it takes concrete form in history in the context-sensitive perception of needs and expectations of free individuals and groups possessed of rights and duties.

It is therefore necessary to promote and facilitate an open debate on the development of robotics which considers reflectively, and critically its intentions, applications and consequences[6]. Such a discussion calls for wide and varied participation which appropriately weighs the differing interests and responsibilities of key actors. The vital contribution of the Christian faith-based perspective to this developing public ethic should not be underestimated.

---

[6] See The European Group on Ethics in Science and New Technologies (2018). *Statement on Artificial Intelligence, Robotics and "Autonomous Systems".*

Conclusion

In view of the complex considerations robotics present for humanity, simple answers are not helpful. There can be no unqualified or emphatic acceptance of these new technologies, nor can there be outright rejection of them, with all their possibilities.

The challenges of scientific and technological development call for a review of the present horizon of principles, a re-examination and re-evaluation of what were previously considered 'settled' norms of behavior and practice. They cause humanity to reconsider its options and priorities in directing individual and social choices, the investment of resources, as well as present and future opportunities.

The primacy of the human person based on the recognition of the human dignity builds the central part of this review. A balanced respect for technological developments and the clear vision for the commitment of the human responsibility to the common good is essential.

It is necessary to be attentive to this developing field of research and innovation and to accompany its actors and processes in a critically reflective, constructive way which seeks to cultivate a public ethic and promote the common good.

This necessitates more than a crude, utilitarian cost-benefit analysis of new technologies in their social, environmental and economic dimensions. It is essential to encourages the development of a humanistic culture which discerns the connections between science and technology and the anthropological, cultural and ethical aspects[7]. Only this multidisciplinary consideration of robotics can help harness the potential of such scientific and technological innovations in ways which respect human dignity and promote the common good.

[7] The term of a 'humanistic culture' has to be shaped by main principles like: rule of law, social justice, solidarity, accountability and transparency.

APPENDIX II

Rome Call for AI Ethics

ABSTRACT

The *Call for AI Ethics* is a document signed by the Pontifical Academy for Life, Microsoft, IBM, FAO and the Ministry of Innovation, a part of the Italian Government in Rome on February 28th, 2020 to promote an ethical approach to artificial intelligence. The idea behind it is to promote a sense of shared responsibility among international organizations, governments, institutions, and the private sector in an effort to create a future in which digital innovation and technological progress grant mankind its centrality. Pointing to a new algorethics, the signatories committed to request the development of an artificial intelligence that serves every person and humanity as a whole; that respects the dignity of the human person, so that every individual can benefit from the advances of technology; and that does not have as its sole goal greater profit or the gradual replacement of people in the workplace.

Introduction

'Artificial intelligence' (AI) is bringing about profound changes in the lives of human beings, and it will continue to do so. AI offers enormous potential when it comes to improving social coexistence and personal well-being, augmenting human capabilities and enabling or facilitating many tasks that can be carried out more efficiently and effectively. However, these results are by no means guaranteed. The transformations currently underway are not just quantitative. Above all, they are qualitative, because they affect the way these tasks are carried out and the way in which we perceive reality and human nature itself, so much so that they can influence our mental and interpersonal habits. New technology must be researched and produced in accordance with criteria that ensure it truly serves the entire "human family" (Preamble, *Universal Declaration of Human Rights*), respecting the inherent dignity of each of its members and all natural environments, and taking into account the needs of those who are most vulnerable. The aim is not only to ensure that no one is excluded, but also to expand those areas of freedom that could be threatened by algorithmic conditioning.

Given the innovative and complex nature of the questions posed by digital transformation, it is essential for all the stakeholders involved to work together and for all the needs affected by AI to be represented. This Call is a step forward with a view to growing with a common understanding and searching for a language and solutions we can share. Based on this, we can acknowledge and accept responsibilities that take into account the entire process of technological innovation, from design through to distribution and use, encouraging real commitment in a range of practical scenarios. In the long term, the values and principles that we are able to instill in AI will help to establish a framework that regulates and acts as a point of reference for digital ethics, guiding our actions and promoting the use of technology to benefit humanity and the environment.

Now more than ever, we must guarantee an outlook in which AI is developed with a focus not on technology, but rather for the good of humanity and of the environment, of our common and shared home and of its human inhabitants, who are inextricably connected. In other words, a vision in which human beings and nature are at the heart of how digital innovation is developed, supported rather than gradually replaced by technologies that behave like rational actors but are in no way human. It is time to begin preparing for more technological future in which machines will have a more important

role in the lives of human beings, but also a future in which it is clear that technological progress affirms the brilliance of the human race and remains dependent on its ethical integrity.

Ethics

All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of fellowship (cf. Art. 1, *Universal Declaration of Human Rights*). This fundamental condition of freedom and dignity must also be protected and guaranteed when producing and using AI systems. This must be done by safeguarding the rights and the freedom of individuals so that they are not discriminated against by algorithms due to their "race, color, sex, language, religion, political or other opinion, national or social origin, property, birth or other status" (Art. 2, *Universal Declaration of Human Rights*).

AI systems must be conceived, designed and implemented to serve and protect human beings and the environment in which they live. This fundamental outlook must translate into a commitment to create living conditions (both social and personal) that allow both groups

and individual members to strive to fully express themselves where possible.

In order for technological advancement to align with true progress for the human race and respect for the planet, it must meet three requirements. It must include every human being, discriminating against no one; it must have the good of humankind and the good of every human being at its heart; finally, it must be mindful of the complex reality of our ecosystem and be characterized by the way in which it cares for and protects the planet (our "common and shared home") with a highly sustainable approach, which also includes the use of artificial intelligence in ensuring sustainable food systems in the future. Furthermore, each person must be aware when he or she is interacting with a machine.

AI-based technology must never be used to exploit people in any way, especially those who are most vulnerable. Instead, it must be used to help people develop their abilities (empowerment/enablement) and to support the planet.

Education

Transforming the world through the innovation of AI means undertaking to build a future for and with younger generations. This undertaking must be reflected in a commitment to education, developing specific curricula that span different disciplines in the humanities, science and technology, and taking responsibility for educating younger generations. This commitment means working to improve the quality of education that young people receive; this must be delivered via methods that are accessible to all, that do not discriminate and that can offer equality of opportunity and treatment. Universal access to education must be achieved through principles of solidarity and fairness.

Access to lifelong learning must be guaranteed also for the elderly, who must be offered the opportunity to access offline services during the digital and technological transition. Moreover, these technologies can prove enormously useful in helping people with disabilities to learn and become more independent: inclusive education therefore also means using AI to support and integrate each and every person, offering help and opportunities for social participation (e.g., remote working for those with limited mobility, technological support for those with cognitive disabilities, etc.).

The impact of the transformations brought about by AI in society, work and education has made it essential to overhaul school curricula in order to make the educational motto "no one left behind" a reality. In

the education sector, reforms are needed in order to establish high and objective standards that can improve individual results. These standards should not be limited to the development of digital skills but should focus instead on making sure that each person can fully express their capabilities and on working for the good of the community, even when there is no personal benefit to be gained from this.

As we design and plan for the society of tomorrow, the use of AI must follow forms of action that are socially oriented, creative, connective, productive, responsible, and capable of having a positive impact on the personal and social life of younger generations. The social and ethical impact of AI must be also at the core of educational activities of AI.

The main aim of this education must be to raise awareness of the opportunities and also the possible critical issues posed by AI from the perspective of social inclusion and individual respect.

Rights

The development of AI in the service of humankind and the planet must be reflected in regulations and principles that protect people – particularly the weak and the underprivileged – and natural environments. The ethical commitment of all the stakeholders involved is a crucial starting point; to make this future a reality, values, principles, and in some cases, legal regulations, are absolutely indispensable in order to support, structure and guide this process.

To develop and implement AI systems that benefit humanity and the planet while acting as tools to build and maintain international peace, the development of AI must go hand in hand with robust digital security measures.

In order for AI to act as a tool for the good of humanity and the planet, we must put the topic of protecting human rights in the digital era at the heart of public debate. The time has come to question whether new forms of automation and algorithmic activity necessitate the development of stronger responsibilities. In particular, it will be essential to consider some form of 'duty of explanation': we must think about making not only the decision-making criteria of AI-based algorithmic agents understandable, but also their purpose and objectives. These devices must be able to offer individuals information on the logic behind the algorithms used to make decisions. This will increase transparency, traceability and responsibility, making the computer-aided decision-making process more valid.

New forms of regulation must be encouraged to promote transparency and compliance with ethical principles, especially for advanced tech-

nologies that have a higher risk of impacting human rights, such as facial recognition.

To achieve these objectives, we must set out from the very beginning of each algorithm's development with an 'algor-ethical' vision, i.e., an approach of ethics by design. Designing and planning AI systems that we can trust involves seeking a consensus among political decision-makers, UN system agencies and other intergovernmental organizations, researchers, the world of academia and representatives of non-governmental organizations regarding the ethical principles that should be built into these technologies. For this reason, the sponsors of the Call express their desire to work together, in this context and at a national and international level, to promote 'algor-ethics', namely the ethical use of AI as defined by the following principles:

1. *Transparency: in principle, AI systems must be explainable.*
2. *Inclusion: the needs of all human beings must be taken into consideration so that everyone can benefit and all individuals can be offered the best possible conditions to express themselves and develop.*
3. *Responsibility: those who design and deploy the use of AI must proceed with responsibility and transparency.*
4 *Impartiality: do not create or act according to bias, thus safeguarding fairness and human dignity.*
5. *Reliability: AI systems must be able to work reliably.*
6. *Security and privacy: AI systems must work securely and respect the privacy of users.*

These principles are fundamental elements of good innovation.

AUTHORS

Francesco Arvizzigno is a Psychologist and passionate about Technology and Automotive.

Tony Belpaeme is Professor at Ghent University and Visiting Professor of Cognitive Systems and Robotics at Plymouth University. He is a member of IDLab-imec at Ghent and is associated with the Centre for Robotics and Neural Systems at Plymouth University.

Tibor Bosse is Chair of the Communication & Media group at Radboud University and President of the Benelux Association for Artificial Intelligence (BNVKI).

Angelo Cangelosi is Professor of Machine Learning and Robotics at the University of Manchester (UK), where he leads the Cognitive Robotics Lab. He also is Turing Fellow at the Alan Turing Institute London, Visiting Professor at Hohai University and at Università Cattolica del Sacro Cuore, Milan, and Turin University, and Visiting Distinguished Fellow at AIST-AIRC Tokyo. His research interests are in developmental robotics, language grounding, human robot-interaction and trust, and robot companions for health and social care. Cangelosi is Editor of the journals *Interaction Studies* and *IET Cognitive Computation and Systems*, and in 2015 was Editor-in-Chief of *IEEE Transactions on Autonomous Development*. His book *Developmental Robotics: From Babies to Robots* (MIT Press) was published in January 2015, and recently translated in Chinese and Japanese. His latest book *Cognitive Robotics* (MIT Press), coedited with Minoru Asada, will be published in 2022.

Alfredo Cesario is Open Innovation Manager for the Scientific Directorate at Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome. Co-founder of the European Association for Systems Medicine (EASyM). Chair of the scientific advisory board of the ERANET 'ERACOSYSMED'. Medical director at Innovation Sprint Sprl.

Daniele Chiffi is Researcher of Logic and Philosophy of Science at the Department of Architecture and Urban Studies of the Politecnico di Milano. He teaches Ethics of Technology and Critical Thinking at the School of Industrial and Information Engineering of the same University. Teaching Fellow at the Bocconi University of Milan. Member of the Steering Committee of the META research group.

Alice Chirico is Research fellow, and co-director of the Experience Lab at the Università Cattolica del Sacro Cuore in Milan, co-Pi of a national project funded by Fondazione Cariplo on the use of theatrical performances focused on the psychological sublime to counteract school dropout in the Italian preadolescent population. She published the first book on deep wonder in Italian and edited the first interdisciplinary text on the link between psychological sublime and major depressive disorder. Editor for the journals *Scientific Reports* (Nature group), *Open Psychology*, *Sustainability and Advances in Human-Computer Interaction*. Her field of study concerns the design and study of complex aesthetic experiences, supported by art and new technologies to promote health and well-being.

Eunae Cho is Data Scientist at SK Holdings C&C, Seoul, South Korea.

Fausto Colombo is Professor of Media Theory and Techniques and Media and Politics at the Università Cattolica del Sacro Cuore, where he is Rector's Delegate for Communication. In 2015, he held the UNESCO Chair in International Communications at the Université Stendhal in Grenoble. He is a member of the Board of ECREA (European Communication Research and Communication Association).

Marika D'Oria is Ph.D. in Education and Communication. Health Communications Officer for the Scientific Directorate at Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome.

Ciro De Florio is Associate Professor of Logic and Philosophy of Science at the Faculty of Economics, Università Cattolica del Sacro Cuore, Milan. He is currently a member of the Scientific Committee of the Humane Technology Lab of the Università Cattolica del Sacro Cuore and a Faculty member of the inter-University doctorate FINO (Philosophy Consortium of the North-West). He was a member of the Italian Society for Analytic Philosophy Steering Committee (2016-2018).

Gabriele Della Morte is Full Professor of International Law at the Law Faculty of the Università Cattolica del Sacro Cuore, Milan. He is a lawyer

admitted on the List of Defence Counsel of the International Criminal Court. He is currently Principal Investigator of a University research project (funding line D.3.2) titled: "Public functions, private control. Interdisciplinary studies on the governance without government of the algorithmic society". He is member of the Scientific Committee of the Humane Technology Lab since its foundation.

Zhong Zhun Deng, School of Management Huazhong University of Science and Technology, Wuhan, China.

Sue Denham is Professor in Cognitive and Computational Neuroscience, a former Director of the Cognition Institute and co-ordinator of Cognitive Innovation (CogNovo), University of Plymouth, UK.

Cinzia Di Dio is Researcher at the Research Unit on the Theory of Mind, Dep of Psychology, Università Cattolica del Sacro Cuore, Milan. She is member of the Doctorate in Sciences of the Person and Education, the International Society for the Study of Behavioural Development (ISSBD), for which she holds the role of Representative of Early Career Scholars, as well as further National and International Societies and Associations primarily dealing with developmental psychology and education.

Daniele Di Lernia is Research Fellow at Humane Technology Lab at Università Cattolica del Sacro Cuore, Milan. Honorary Associate Researcher at Royal Holloway University of London. Researcher at Applied Technology for Neuro-Psychology Laboratory at the Istituto Auxologico Italiano.

Fabio Fossa is a Research Fellow at the Department of Mechanics of the Politecnico di Milano, where he deals with the philosophy of artificial agents and ethics of autonomous vehicles. He is director of the magazine *InCircolo - Journal of philosophy and culture*, a member of the steering committee of the META group, and a founding member of the Zetesis research group.

Jesse Fox is Associate Professor in the School of Communication and Director of the Virtual Environment, Communication Technology, and Online Research (VECTOR) Lab, The Ohio State University, USA.

Andrea Gaggioli is Full Professor of General Psychology at the Faculty of Humanities, Università Cattolica del Sacro Cuore, Milan. He directs the International Master of User Experience Psychology and the Research Unit of Psychology of Creativity and Innovation.

Andrew Gambino is Scholar of communication, technology, and relationships, Bellisario College of Communications, USA.

Matteo Gatti is Associate Professor at the Faculty of Agricultural Food and Environmental Sciences of the Università Cattolica del Sacro Cuore, Piacenza. Professor of Viticulture, he deals with precision agriculture and automation within national and international university programs. Referent for the Università Cattolica del Sacro Cuore of the joint laboratory for robotics in agriculture involving Università Cattolica del Sacro Cuore and Italian Institute of Technology.

Teresa Giovanazzi is Research Fellow at the Faculty of Education at the Free University of Bozen-Bolzano. She collaborates with the Graduate School for Environment of the Università Cattolica del Sacro Cuore on the issues of sustainability and educational planning with particular reference to universal exhibitions.

Guendalina Graffigna is Professor of Psychology of Consumer and Health at the Faculty of Agricultural, Food and Environmental Sciences and Director of the EngageMinds HUB - Consumer, Food & Health Engagement Research Center at Università Cattolica del Sacro Cuore, Milan.

Paolo Guadagna is Research Fellow at the Department of Plant and Sustainable Production of the Università Cattolica del Sacro Cuore, Piacenza. He works in the field of robotics in viticulture.

Michaela Gummerum is Associate Professor, Department of Psychology, University of Warwick, UK.

Yaniv Hanoch is Associate Professor in Risk Management, Southampton Business School, UK.

Daniel Hernandez García is Postdoctoral Research Associate at the Interaction Lab in the School of Mathematical and Computer Sciences, Heriot-Watt University, UK.

Evelien Heyselaar is Postdoctoral Researcher at the Communication & Media, Behavioural Science Institute and Assistant Professor at Communication Science Department, Radboud University Nijmegen, The Netherlands.

Sara Incao is Ph.D. student in Bioengineering and Robotics at the University of Genoa. She carries out her research at the Italian Institute of

Technology in the department of Cognitive Architecture for Collaborative Technologies (CONTACT). She graduated in Philosophy at the Università Cattolica del Sacro Cuore, Milan. Her research currently concerns the idea of Self in robots.

Hiroshi Ishiguro is Professor of Department of Systems Innovation in the Graduate School of Engineering Science at Osaka University (2009-) and Distinguished Professor of Osaka University (2017-). He is also visiting Director (2014-) (group leader: 2002-2013) of Hiroshi Ishiguro Laboratories at the Advanced Telecommunications Research Institute and an ATR fellow.

Mitsuhiko Ishikawa is JSPS Research Fellow of Centre for Baby Science, Doshisha University. Visiting Research Fellow of Centre for Brain and Cognitive Development, Birkbeck, University of London. His research focuses on how humans perceive social partner's gaze information (e.g., direct gaze, gaze cueing) and respond to it (e.g., gaze following). He was awarded Kyoto University President's Award (2020) and JSPS Ikushi Prize (2021).

Shoji Itakura is Director of the Center for Baby Science, Doshisha University, Fellow Professor, President of the Japanese Society of Baby Science, Emeritus Professor of Kyoto University, and Visiting Professor at Università Cattolica del Sacro Cuore, Milan, Zhejiang Normal University, Zhejiang Sci-Tech University.

Yoonhyuk Jung is Associate Professor in the School of Media & Communication at Korea University in South Korea. He is currently the vice-president of the Korean Association for Information Society.

Takayuki Kanda is Professor in Informatics at Kyoto University, Japan. He is also a Visiting Group Leader at ATR Intelligent Robotics and Communication Laboratories, Kyoto, Japan.

Seongcheol Kim is Professor in the School of Media and Communication at Korea University and the Korea University Librarian. He is the founder and director of the Center for Media Industry (CMI). He is also the former president of Korea Media Management Association (KMMA) and the vice president of the Korean Association for Information Society (KAIS).

Mario A. Maggioni is Professor of Political Economy at the Faculty of Political and Social Sciences and Head of the Department of Internation-

al Economics, Institutions and Development (DISEIS) at Università Cattolica del Sacro Cuore, Milan. In the same University, he is also the Director of CSCC (Research Center on Cognitive Science and Communication) and HuRo-Lab (Laboratory on Human-Robot interactions), where techniques of behavioral economics are applied to the study of Human-Robot interactions.

Pierluigi Malavasi is Professor of General and Social Pedagogy at the Università Cattolica del Sacro Cuore, Milan. He is the coordinator of the Master in Pedagogical Design and Training of Human Resources. He is delegate of the Rector in the Management Committee of the Graduate School for Environment. He is the scientific director of the first level Master in Environmental Governance for Integral Ecology. Climate risk, adaptation, training.

Clelia Malighetti is Ph.D. student at the Faculty of Psychology of the Università Cattolica del Sacro Cuore, Milan.

Fabrizia Mantovani is Professor of General Psychology, Department of Human Sciences for Education "Riccardo Massa", University of Milan Bicocca.

Federico Manzi is Researcher in Developmental Psychology and Educational Psychology, Faculty of Education, Università Cattolica del Sacro Cuore, Milan. He is a member of the Research Unit on Theory of Mind, Department of Psychology, Università Cattolica del Sacro Cuore, Milan. He is an adjunct researcher at the Institute of Psychology and Education, University of Neuchâtel. He is also member of the Early Career Committee of the International Society of Behavioural Development (ISSBD). He is tutor of the Specialization Course in "Growing and living with robots: psychology, education, and care at the time of social robots. Design and development of robot-assisted interventions". He is member of the Expert Advisory Board of the EU Horizon Project "Robotics4eu".

Massimo Marassi is Professor of Theoretical Philosophy at the Faculty of Humanities of the Università Cattolica del Sacro Cuore, Milan. He directs the Department of Philosophy, the Master in Philosophical Expertise for Economic Decisions, the *Rivista di Filosofia Neo-Scolastica*. He is the coordinator of the Quality Assurance Unit of the research activities – PQA – of the University.

Antonella Marchetti is Professor of Developmental Psychology and Educational Psychology, Università Cattolica del Sacro Cuore, Milan. Head of

the Department of Psychology, Director of the Research Unit on Theory of Mind, Coordinator of the Doctorate in Sciences of the Person and Education, Member of the Scientific Board of the Humane Technology Lab of the same University. She is ISSBD Executive Committee member and ISSBD Regional Coordinator for Italy and member of the Scientific Committee of FEDUF (Foundation for the Financial Education and Saving). She co-directs the Specialization course in "Growing and living with robots: psychology, education and care at the time of social robots. Design and development of robot-assisted interventions" (Crescere e vivere con i robot: psicologia, educazione e cura al tempo dei robot sociali. Progettazione e sviluppo di interventi robot-assistiti).

Giovanna Mascheroni is Associate Professor of Sociology of Digital Media, Faculty of Political and Social Sciences, Università Cattolica del Sacro Cuore. He coordinates the "WP6 of the Horizon ySKILLS" project. Principal Investigator of the DataChildFutures project – Fondazione Cariplo Call for proposals Social Research 2019 – Science, Technology and Society.

Davide Massaro is Professor of Developmental Psychology and Educational Psychology, Università Cattolica del Sacro Cuore, Milan. He is a 'senior' member of the Research Unit on Theory of Mind; Coordinator of the Research Unit of Psychology of Religion; Didactic Coordinator of the Doctorate in Sciences of the Person and Education and co-coordinator of the Master's degree course in Media Education. He co-directs the Specialization course in "Growing and living with robots: psychology, education, and care at the time of social robots. Design and development of robot-assisted interventions" (Crescere e vivere con i robot: psicologia, educazione e cura al tempo dei robot sociali. Progettazione e sviluppo di interventi robot-assistiti).

Carlo Mazzola is Ph.D. student in Bioengineering and Robotics at the University of Genoa. He carries out his research at the Italian Institute of Technology in the department of Robotics Brain and Cognitive Sciences (RBCS) on Shared Perception between humans and robots. Previously, he graduated in Theoretical Philosophy at the Università Cattolica del Sacro Cuore, Milan.

Barbara C.N. Müller is Assistant Professor at the Communication Science department, Radboud University Nijmegen, The Netherlands.

Sari R.R. Nijssen is Ph.D. candidate, Behavioural Science Institute, Radboud University Nijmegen, The Netherlands.

Giulia Peretti is Ph.D. student at the Faculty of Education at the Università Cattolica del Sacro Cuore, Milan. She is a member of the Research Unit on Theory of Mind, Department of Psychology, at the same University.

Stefano Poni is Professor of Viticulture at the Faculty of Agricultural, Food and Environmental Sciences of the Università Cattolica del Sacro Cuore, Piacenza. Director of the First Level International Master VENIT (Viticulture and Enology: Innovation meets tradition), Coordinator of the Master of Science in Sustainable Viticulture and Enology, and Chair of the Value Chain SOS-FARM (Sustainable and Precision Agriculture) for the Emilia Romagna Region.

Francesco Rea is Researcher at the Italian Institute of Technology, Robotics, Brain and Cognitive Sciences department, is research leader of the cognitive interaction lab and also responsible for IIT of three European projects on collaborative robotics: H2020 APRIL, H2020 VOJEXT and HBP PROMEN-AID.

Giuseppe Riva is Professor of General Psychology at the Faculty of Psychology, Università Cattolica del Sacro Cuore, Milan. He directs the Humane Technology Lab of the Università Cattolica del Sacro Cuore. He directs the Applied Technology for Neuro-Psychology Lab of the Istituto Auxologico Italiano. President of the International Association of CyberPsychology, Training, and Rehabilitation.

Pier Cesare Rivoltella is Professor of Technologies of Education and Learning at the Faculty of Education at Università Cattolica del Sacro Cuore, Milan. He directs the CREMIT (Research Center on Media Education, Innovation and Technology). Member of the School Commission of the Accademia dei Lincei, he is the President of SIREM (Italian Society for Research on Media Education).

Domenico Rossignoli is Researcher of Economic Policy at the Faculty of Political and Social Sciences of the Università Cattolica del Sacro Cuore, Milan. At the same university, he is a member of the HuRo Lab, the Research Center in Cognitive and Communication Sciences (CSCC), and the Department of International Economics, Institutions and Development (DISEIS).

Giulio Sandini is Founding Director of the Italian Institute of Technology and coordinator of the Robotics, Brain and Cognitive Sciences Unit. He was Assistant Professor at the Scuola Normale Superiore of Pisa, Visiting Researcher at the Neurology Department of the Harvard Medical

School and Full Professor of Bioengineering at the University of Genoa. His research activity is characterized by an engineering approach to the study of natural intelligent systems with a focus on the design and implementation of artificial systems to investigate the development of human perceptual, motor and cognitive abilities (and viceversa).

Alessandra Sciutti is Researcher in charge of CONTACT – Cognitive Architecture for Collaborative Technologies unit at the Italian Institute of Technology. She holds a degree in Bioengineering and a Ph.D. in Humanoid Technologies from the University of Genoa. After two research periods in the USA and Japan, in 2018 she received a prestigious European ERC-Starting Grant for young researchers with the wHiSPER project, aimed at establishing a mutual understanding between humans and robots. Her research is aimed at studying the sensory, motor and cognitive mechanisms underlying human interaction, with the technological goal of designing robots capable of interacting naturally with us humans.

Angela Sorgente is Research fellow in Psychometrics at Università Cattolica del Sacro Cuore, Milan, Italy. Research interests: quantitative methods and emerging adults' well-being.

Simone Tosoni is Associate Professor at the Department of Communication and Performing Arts at Università Cattolica del Sacro Cuore, Milan, where he teaches courses on the sociology of cultural processes and digital media. He is part of the ECREA (European Communication Research and Education Association) board and lecturer at the ECREA Doctoral Summer School. He is currently working on media machines, social robotics, and the online circulation of knowledge refused by the scientific community. Among his publications, *Entanglements. Conversations on the Human Traces of Science, Technology, and Sound* (MIT Press, 2017, with Trevor Pinch).

Vincenzo Valentini is Professor of Oncological Radiotherapy at the Faculty of Medicine and Surgery of the Università Cattolica del Sacro Cuore, Rome. Director of the Department of Diagnostic Imaging, Oncological Radiotherapy and Hematology at Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome. Deputy Director for 'Big Data' for the Scientific Directorate at Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome. Scientific Director of the Gemelli Generator Laboratory at Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome.

Daniela Villani is Associate Professor of General Psychology at the Facul-

ty of Education of the Università Cattolica del Sacro Cuore, Milan, and head of the Research Unit of Digital Media, Psychology and Wellbeing at the same university.

Brenda K. Wiederhold is President of the Virtual Reality Medical Center (VRMC). She is Chief Executive Officer of the Interactive Media Institute and Visiting Professor at the Università Cattolica del Sacro Cuore, Milan, and an Advisory Board Member for the International Child Art Foundation. Founder of the international CyberPsychology, CyberTherapy, & Social Networking Conference (CYPSY) and Editor-in-Chief of the *CyberPsychology, Behavior, and Social Networking Journal* (CYBER).

Dong Hong Zhu is Associate Professor, School of Management Huazhong University of Science and Technology, Wuhan, China.