From social brains to social robots: Applying neurocognitive insights to human-robot interaction

Emily S. Cross[1], Ruud Hortensius[1], Agnieszka Wykowska[2]

[1] *Institute of Neuroscience and Psychology, School of Psychology, University of Glasgow, Scotland*

[2] *Istituto Italiano di Tecnologia, Genoa, Italy*

All authors contributed equally to this paper. Authorship is presented alphabetically.

Contact:

E.S. Cross: emily.cross@glasgow.ac.uk

R. Hortensius: ruud.hortensius@glasgow.ac.uk

A. Wykowska: Agnieszka.Wykowska@iit.it

Accepted version of Cross ES, Hortensius R, Wykowska A. 2019 From social brains to social robots: applying neurocognitive insights to human–robot interaction. Phil. Trans. R. Soc. B 20180024. http://dx.doi.org/10.1098/rstb.2018.0024

Abstract

Amidst the fourth industrial revolution, social robots are resolutely moving from fiction to reality. With sophisticated artificial agents becoming ever more ubiquitous in daily life, scientists across different fields are grappling with the questions concerning how humans perceive and interact with these agents and the extent to which the human brain incorporates intelligent machines into our social milieu. This theme issue surveys and discusses the latest findings, current challenges and future directions in neuroscience- and psychology-inspired human-robot interaction (HRI). Critical questions are explored from a transdisciplinary perspective centred around four core topics in HRI: technical solutions for human-robot interaction, development and learning for human-robot interaction, robots as a tool to study social cognition, and moral and ethical implications of human-robot interaction. Integrating findings from diverse but complementary research fields, including social and cognitive neurosciences, psychology, artificial intelligence, virtual reality and robotics, the contributions showcase ways in which research from disciplines spanning biological sciences, social sciences and technology deepen our understanding of the potential and limits of robotic agents in human social life.

1. Introduction

As artificial intelligence and engineering technology continues to develop, we are encountering ever more sophisticated "social" robots – not just in films and television series, but increasingly in real world contexts [1]. The prospect of interacting with these "social robots" in our daily lives is on the horizon, and yet, remarkably little is known about how we perceive, interact with or accept these machines in social contexts [2,3]. Underscoring the importance of this anticipated infiltration of machines in human society, major research agencies around the world, including the European Commission and the National Science Foundation, aim to strengthen their role in supporting robotics development, articulating strategic visions for the integration of robots into every part of society. Similarly, multinational companies including Facebook, Amazon and Google, as well as countless smaller start-ups and medium sized tech firms, are continuing to invest significant sums in the development of artificial intelligence and robotics [4]. These examples and many others forecast the growing presence of artificial agents in social environments.

As a consequence, important questions are emerging regarding the flexibility and adaptability of human social cognition when perceiving, communicating with or cooperating with artificial agents. Research interest from experimental psychology, social neuroscience, computer science and robotics exploring how humans interact with artificial agents has been making progress alongside the technological advances [2,3,5-8]. However, in order for significant advances to be made in understanding how human social cognition interfaces with artificial actors, deeper and more constructive dialogue and collaboration across these domains is required. One area that holds exceptional potential along these lines is the application of rigorous experimental methods from cognitive/social neuroscience and psychology to studies examining how we perceive and interact with artificial agents. This burgeoning research area stands to not only construct a more sophisticated understanding of the neurocognitive mechanisms and consequences of human-artificial agent interactions, but also promises to inform the development of increasingly socially sophisticated robots.

This theme issue captures the most cutting-edge scientific visions and findings that drive this emerging field. While previous papers published individually have focussed on one specific aspect of the challenges or opportunities encountered when human minds interface with socially intelligent technology, this theme issue aims to provide a systematic, multi-level compendium of the impact of interaction with this technology on human perception, cognition and behaviour. By fostering a constructive and critical dialogue between the diverse disciplines that explore the social and cognitive neuroscience of human-robot interaction, this theme issue presents an integrated body of knowledge and overview of the state of the art within the relevant disciplines, which stands to not only advance our understanding of human-robot relations, but also articulate a transdisciplinary agenda for future research. On a more global level, through this theme issue we aim to showcase a framework to describe and examine human-robot interactions at different levels, from basic to complex, across behavioural and brain levels, from the human side as well as from the perspective of robot social capabilities.

The contributors to this theme issue draw upon different disciplinary expertise to address novel questions at the core of establishing the potential and limits of human social relationships with robotics technology. The contributions fit within four broad categories, ranging from technical solutions to human-robot interaction (HRI) to learning and development perspectives on social robotics, to contributions underscoring the utility of social robotics research to advancing our knowledge of fundamental aspects of human social cognition, to pieces considering the moral and ethical implications of the (social) relationship between humans and technology. The specific questions addressed by each contribution are highlighted in the following section themes.

2. Technical solutions for human-robot interaction

The first section of the theme issue is dedicated to technical solutions for human-robot interaction. If robots are to be introduced to human environments, they need to be endowed with interaction capabilities. This requires an enormous effort in the fields of engineering and artificial intelligence and covers areas such as face and emotion recognition, action and intention prediction, speech processing and many other. Indeed, social robots need to be able to sense signals from humans, respond adequately, understand and generate natural language, have reasoning capacities, plan actions and execute movements in line with what is required by the specific context or situation. In this part of our theme issue, three contributions cover the areas of research related to technical solutions for human-robot interaction: computational architectures (Prescott and colleagues), classification and prediction of human behaviour and expressions (Gunes and colleagues), and natural language processing (Foster). These three contributions provide examples of challenges that roboticists and artificial intelligence experts need to face in order to design robots endowed with capabilities crucial for social interactions with humans.

The opening paper of this section is by Prescott and colleagues [REF] who propose a neuroscience-inspired multimodal computational architecture for autobiographical memory system – the Mental Time Travel (MTT) Model – implemented on the iCub robot [9]. The model allows for retrieving past events, and project into imagined future by using the same system. This architecture proves useful for social capabilities of robots by enabling face, voice (including emotion), action and touch gesture recognition through interaction with humans. Using this system for imagining future events should allow for simulating and visualizing actions, as well as planning actions before actual execution. Apart from clear usefulness of such a model for human-robot interaction, this approach has also a potential for contributing to more fundamental and theoretical research questions. In line with authors' argumentation, this approach shows that when computational models are implemented in an embodied agent, one can test whether the models correctly capture cognitive mechanisms through examining generated behaviour, and interaction with the environment.

This paper is followed by a contribution by Gunes and colleagues [REF] which is focused on the capability of artificial systems for online, real-time automatic prediction of personality of human users, based on their behavioural cues, and emotional facial expression. The authors report a method for user's personality prediction based on features such as facial appearance, geometric facial and body features, as well as temporal relations between the extracted features and continuous annotations. The paper describes implementation of such methods in a NAO robot (Softbank Robotics) during a number of public demonstrations at various exhibitions and international conferences.

Finally, Foster [REF] addresses another crucial competence that is required to enable natural human-robot interaction: the ability of the technical system to understand spoken natural language and respond appropriately. Foster provides an overview of methods used in human-robot interaction for natural language generation. In this opinion piece, she points out that methods for natural language generation are mainly developed in the computational linguistics community, while the field of social robotics, which needs to address several other technological challenges, has so far devoted less space for the development of natural language generation methods, relying often on simple and traditional template-based or rule-based approaches. Interestingly, however, as Foster highlights, the area of social robotics is one of the most suitable testbeds for situated language generation research, where more recent data-driven approaches could be exploited. She suggests that social robotics would benefit for incorporating data-driven and sophisticated machine learning techniques in human-robot interaction scenarios, considering particular contexts in which the interaction takes place.

3. Development and learning for human-robot interaction

The second section deals with the development of social behaviour in robots. For adequate human-robot interaction, the robotic agent needs to not only show sophisticated ways of problem solving, but also develop social engaging and relevant behaviours. Ultimately, humans will interact with a completely autonomous, social agent that continuously develops and learns new social behaviours and adapts to new social challenges. The contributions in this section are at the core of developmental robotics [10], thereby bridging cognitive robotics and social robotics, and work towards achieving autonomous behaviour in robots by embracing a bottom-up embodied cognition and internal learning approach. As the contributions show, these biology and psychology-inspired social robotics systems can lead to an impressive range of social engaging behaviours, such as attention-grabbing emotional facial expressions (Gordon), and even trust and theory-of-mind (Cangelosi). Crucially, these perspectives directly provide new insights for developmental psychology, for example into developmental mechanisms of social behaviour as well as clinical insights into autism spectrum condition (Kuniyoshi and Nagai). Together, these contributions move away from scripted, fully top-down controlled robotic system as often seen in demos of robotic systems, and move toward fully autonomous social agents that are inspired and build upon core concepts from human psychology and neuroscience.

In the first contribution of this section, Kuniyoshi [REF] provides an integrative review of biology-inspired models of early development that are foundational for social interaction. In this framework, embodied cognition, the interaction between the agent and the environment, is crucial for the emergence of autonomous social behaviour and continuous development of these behaviours. After outlining a theoretical model for early development in the context of social robots, Kuniyoshi presents empirical evidence on the emergence of simple and complex behaviour as a consequence of embodied sensorimotor interactions in both a simple robotic system and a complex human foetal body model. Independent of external reward or learning, adaptive behaviours occur ranging from locomotion in a robot to the emergence of a body schema and multi-modal sensory integration in the human foetal body model. While these findings already provide fundamental insight into the developmental of autonomous robotic systems, they provide a framework and model to systematically test outstanding questions on human development. For instance, Kuniyoshi highlights the potential by studying the impact of preterm birth on the development of body schema and multisensory integration in the human foetal body model. These results have direct implications for conditions that showcase impairments in these processes such as autism spectrum condition.

The second contribution in this section by Gordon [REF], continues to highlight the importance of the interplay between developmental robotics and psychology, and the crucial role of the embodiment of the agent in the environment, to develop autonomous social robots. In this opinion piece, the author argues that the emergence of social behaviour is the result of an interaction between the embodiment of the agent in the environment and the curiosity drive of agent. This curiosity drive is the intrinsic motivation of an agent to maximise learning about themselves and the environment. Again, as seen in the contribution by Kuniyoshi, this bottom-up process can lead to simple to complex social behaviour dependent on the type of environment and agent. In a non-social environment, artificial curiosity in an embodied physical agent can lead to the emergence of motion detection and even self-awareness and infant-like exploration behaviours, while in a social environment this can result in a variety of social behaviours such as seeking the attention of agents in the environment by making attention grabbing facial expressions. Crucially, none of these behaviours are pre-programmed, but are emergent properties of artificial curiosity within a physical agent.

In the third contribution, Nagai [REF] proposes that predictive learning of sensorimotor signals plays a key role in early social cognitive development. Predictive learning is the process of minimising prediction errors between the internal top-down model of the sensorimotor system and actual sensorimotor signals. The author proposes a computational theory in which two interlinked mechanisms to minimise prediction errors support cognitive development. A first mechanism updates a predictor based on experience, while the second mechanisms executes an action anticipated by the predictor. The latter occurs in social situations to minimise the confusion between other individual's actions and the prediction of internal states. When employed in robots, such as the iCub robot, the first mechanism or experience-dependent updating of predictors, leads to self-other cognition and goal-directed actions. The second mechanism underpins social behaviours such as imitation and helping behaviours. Converging with previous contributions in this section, the review show how implementation of psychology-driven algorithms in robots can replicate developmental dynamics observed in infants and provide further evidence for theories of human developmental [11].

In the final contribution of this section, Vinanzi and colleagues [REF] ask the intriguing question if and how robots can trust humans. Trust is foundational for collaborations between two or more agents, and so far, research on human-robot interaction has mainly looked at this from the perspective of the human agent. The authors first present the artificial cognitive architecture required for trust behaviour, and subsequently provide empirical confirmation of this behaviour in a humanoid Pepper robot (Softbank Robotics) and compare these results with a developmental sample. In order to establish trustworthiness of the human agent, the authors argue that Theory of Mind is crucial. Theory of Mind is the ability to attribute intentions and other mental states to other agents. Pairing a probabilistic Theory of Mind with a trust model supported by episodic memory, a system was developed that allows a robot to not only judge the trustworthiness of a human in a current interaction, but also that of a human with whom the robot never interacted with. This architecture was tested in a decision-making task where the robot has to distinguish between helpful or not helpful intentions of the human agent. Matching results obtained in children, establishment of the trustworthiness of the human agent was dependent upon the presence of a Theory of Mind in the robot. Similarly, if the robot was repeatedly paired with a non-helpful human partner, the robot will distrust new human agents more. Together, these results show the potential of a developmental robotics approach to achieve complex autonomous social behaviour in robots, such as trust behaviour.

4. Robots as a tool to study social cognition

The third thematic section in this issue highlights work by researchers who primarily work within experimental psychology and/or human neuroscience, and who use robots as a specialised tool to address fundamental question about human social perception and social behaviour. The contributions in this section are primarily empirical studies that showcase an eclectic mix of some of the cutting-edge research at the frontier where behavioural and brain sciences meet technology.

This section begins with an elegant study by Rauchbauer and colleagues [REF], who combine functional magnetic resonance imaging (fMRI) with a real-time interaction paradigm wherein participants carried out a conversation (describing photographs) with either a human or robotic interaction partner. The authors aimed to develop a paradigm that maximised the ecological validity of the social interaction task when conversing with both the human and robotic interaction partner. The brain imaging findings revealed that performing the task with a human interaction partner was associated with brain regions associated with higher-order social cognitive processes, including the temporo-parietal junction, while performing the same task with a robotic interaction partner activated dorsal frontal and parietal brain regions. The authors suggest this pattern of findings is indicative of human interactions engaging more social motivation and mentalising, while interactions with robots recruit additional executive and perceptual resources, and emphasize the utility of this rich, interactive paradigm for studying the neurocognitive consequences of human-robot interaction.

The second piece in this section is also an fMRI study, conducted by Cross and colleagues [REF]. In this ambitious training study, the authors were interested in the effects of socialising with a socially engaging robot (the palm-sized Cozmo robot by Anki, Inc) on participants' empathy for the robot's emotions (compared to a human's emotions). To do this, the researchers asked a group of participants to take part in two identical brain scanning sessions wherein they watched videos of a human actor and the Cozmo robot receive either pleasant or painful electrical stimulation, and in between the fMRI sessions, participants were given a Cozmo robot to take home and interact with daily over a one-week period. While the use of a repetition suppression design and functional localisers for the pain network allowed the authors to sensitively probe neural representations for pleasure and pain that might be shared between human and robotic agents, the findings overall suggested that the short socialisation intervention with a robot does not demonstrably increase empathy toward the robot, nor does it shape neural responses to look more human-like.

In the third piece, Willemse and Wykowska [REF] worked with the iCub humanoid robot to explore questions related to quantifying the reward value of joint attention with a robot. Participants' primary task was to look at one of two objects presented on two different screens while sat opposite the iCub robot. In a clever manipulation of robot sociality, authors introduced the robot as either "Jimmy" (a pro-social robot who followed participants' gaze 80% of the time) or "Dylan" (a less social robot who only followed participants' gaze 20% of the time). The authors found that participants were quicker to return their gaze to Jimmy, the prosocial robot, after trials when Jimmy followed their gaze, while no such relationship emerged from Dylan, the less social robot. Participants also reported liking Jimmy better. The authors conclude with important considerations of the utility of embodied social interactions with robots that maximise experimental control and ecological validity.

The penultimate piece in this section is a transcranial direct current stimulation (tDCS) empirical piece by Wiese and colleagues [REF], which aimed to examine the role played by key brain regions involved in social perception when an observed agent is perceived as having a mind (human) or not (robot). The authors used a social attention gaze cueing paradigm featuring a human or a robot face, while tDCS was applied over prefrontal and temporo-parietal brain regions. The authors did not find evidence that the temporo-parietal stimulation influenced mechanisms of social attention for either the human or robotic agent, while prefrontal stimulation was associated with increased attentional orienting in response to human gaze cues and decreased attentional orienting in response to the robotic gaze cues. The authors conclude by suggesting that their results support the notion that in this kind of task, mind perception modulates lower-level mechanisms of social cognition via prefrontal structures, while higher-level mechanisms, such as those thought to be modulated by temporo-parietal structures, are less involved.

The final piece for this section is the paper by Collins [REF] which provides a comparative approach for analysing human-robot interaction, by referring to human-animal interaction. Collins points out that human-robot interaction is a system which consists of three elements: the human, the robot and their interaction. Each element of the system needs to be analysed in-depth. In terms of the robot system, its morphology and specific purpose in a given human-robot interaction setup are crucial. In terms of the human element, user's expectations and attitudes with respect to the robot need to be analysed. Finally, the interaction element should be analysed from the perspective of its aims and type. Following a comparative approach, Collins argues that analytical tools for human-robot interaction can be inspired by methodologies used in other domains, such as human-animal interaction. Collins refers to research on social attachments, bonds and relationships that humans form with other humans, animals or objects and reports a study in which she examined the impact of interaction with a biomimetic robot Paro (PARO Robotics) on felt security, and the role of individual differences on the type of social engagement (tactile contact). The paper concludes with highlighting the need to cross boundaries between disciplines, and draw inspiration, for example, from theoretical and methodological basis in social-cognitive psychology for human-robot interaction studies.

5. Moral and ethical implications of human-robot interaction

The fourth and final section of this issue deals with consequences of human-robot interaction in the moral and ethical realm. The contributions discuss implications of increasingly sophisticated and ubiquitous social technology from diverse perspectives, including anthropomorphism, dehumanisation, and consequences of human embodiment in robots, taking into account findings and perspectives from social psychology, philosophy and cognitive neuroscience. Consideration of moral and ethical implications alongside technical and applied aspects of social robotics should inform public and political discourse about the growing role of robots in tomorrow's society.

In the first contribution of this section, Skewes, Seibt, and Amodio [REF] present new and exciting theoretical ideas on the use of social robots as a research tool in the study of biases of social perception and behaviour. In this opinion piece, they present the notion of Fair Proxy Communication, a form of communication in which a human is embodied in a robot with tight control of the communicated information on the identity (e.g. gender) or behaviour (e.g. expressions) of the human. This form of tele-presence allows to study the parameters and mechanisms that lead to biases in social perception, as one can tightly control the form, movement and other features of the robot. They further argue that Fair Proxy Communication can be used to effectively prevent or counteract biases in perception such as stereotypes or favouring ingroup members. For example, no information on the identity or behaviour of a human embodied in the Telenoid R1 robot [12] enters the equation during an interaction (e.g. job interview), preventing the potential impact of stereotypes on the outcome of this interaction. The ideas presented in this contribution provide, as the authors suggest, exciting new implications for social justice with ultimately positive ethical implications of human-robot interaction.

In the second contribution, Wullenkord and Eyssel [REF] set out to test if imagined contact with a robot would improve later interaction with a robot. Some people experience negative attitudes toward robots, as well as anxiety to interact with a robot. The authors borrowed insights from social psychology, where imagined contact, or the simulation of an interaction with a member of an outgroup, reduces negative attitudes and improves behaviour during an interaction with a member of this outgroup. Participants briefly imagined a scenario that was either similar or dissimilar to the future interaction with a small humanoid NAO robot (Softbank Robotics). As predicted, an imagined contact task similar to the subsequent HRI resulted in higher self-rated quality of the interaction, as well as more positive social behaviours toward the robot, compared to an imagined contact task that was dissimilar to the subsequent HRI. These results that imagined contact is an effective method that can be used to change implicit and explicit attitudes and behaviours toward robots.

Waytz and Young [REF] continue this section with a unique and systematic empirical investigation of people's aversion to playing God, which they define as "taking on the role of some higher, metaphysical power to intervene in natural or human affairs". This topic on the moral implications of technology is important with the increasing reliance on novel and sophisticated technologies in everyday life. Across seven studies the authors provide evidence that the aversion to play God predicts negative moral judgements of scientific and technological practices, ranging from genetically modified organisms, to stem cell research, and robotics. For instance, the more people deem the use of autonomous warfare drones as an act of playing God, the lower they rated the moral acceptability of the use of autonomous warfare drones. As the authors rightfully argue, advances made in robotics, artificial intelligence, and other fields of science and technology, warrant further consideration of these questions and the impact on public opinion and acceptance of technology.

Changing gears, Powell and Michael [REF] discuss people's commitment to robots from a cognitive philosophical perspective in the final piece in this section. Commitment is crucial to consider in short-term and long-term human-robot interaction, and the authors centre their opinion piece on the why, what, when and how of human commitment to a robot. The authors argue that the sense of commitment is equally important for human-robot interaction as trust, and sketch how a sense of commitment is dependent on the agent's motivation, expectation, and beliefs as well as cues expressed by the partner, and situational factors. After identifying situations in which a sense of commitment is important given a higher likelihood of disengagement due to an over- or underperforming robot, the paper highlights ways to achieve a sense of commitment during human-robot interaction. Besides explicit verbal communication, the authors discuss implicit forms of commitment such as cognitive effort and physical effort (e.g., direction and speed of a movement). Importantly, these considerations provide direct ways to target the human *and* robot sides of the human-robot interaction.

6. General Discussion

As this theme issue aims to make abundantly clear, research into HRI requires expertise and contributions from a diverse range of scholars and disciplines and calls for continuous dialogue across different fields. The collection of papers in this theme issue illustrates the breadth of research areas that need to be integrated in order to make meaningful progress in HRI. The papers included here cover the topics of technological solutions for human-robot interaction, computational approaches, social and cognitive psychology, social neuroscience, developmental psychology and developmental robotics, as well as ethics. And even *within* each of these topics, the need for interdisciplinary approach has been highlighted. For example, in order to build computational models of social cognition that could be implemented on a social robot, social neuroscience needs to be integrated with AI and robotics. For developing natural language processing for HRI, computational linguistics must be linked with social robotics. Psychological research needs to develop closer bonds with AI so that methods for social signal processing can be developed. We also must emphasize that what is covered in this theme issue is by no means exhaustive, but instead highlights a subset of areas, between which meaningful dialogue and collaboration will help to accelerate progress in HRI.

As many researchers can attest, interdisciplinary enterprises are not an easy task. The challenge lies in establishing a proper methodology that draws inspiration from the contributing fields, but that does not fall in the trap of insufficient scrutiny. The other challenge is in establishing a common scientific or scholarly language that allows all parties to follow and contribute to the discourse. Finally, and more pragmatically, an enormous challenge lies in establishing common platforms for exchange of knowledge and ideas for future directions in research. With this in mind, we developed the idea for this interdisciplinary theme issue and our hope is that it will provide a useful basis for future dialogue between experts in AI, robotics, psychology, linguistics, cognitive and social neuroscience and other relevant areas.

We would also like to emphasize the role that open scientific practices should play in future work in this exciting and rapidly progressing research domain (and which some contributions in this theme issue already show evidence of embracing). In psychology and neuroscience (and clearly many, if not all, other research domains that rely upon empirical research that involve humans as researchers), an abundance of studies with low power and a publication bias skewed towards positive results has produced a poor level of evidence for many claims made [13,14]. Similar issues afflict the fields of computer science [15], and artificial intelligence [16]. Low reproducibility is a problem for the accumulation of knowledge, as well as an inefficient use of public funds. As such, improving the efficiency and robustness of scientific research has important societal impact. Low reproducibility in scientific studies is the result of a complex and system-wide problem, which requires a diverse set of solutions [17,18], including pre-registration of research questions, materials, and hypotheses, power analyses, replication, meta-analyses, open data, and sharing of pre-prints prior to publication. Enthusiasm for open science practices is sweeping across the behavioural and brain sciences, and we strongly urge that individuals working in social robotics adopt these suggestions in their research practices as well, to ensure we are all working together to build a reliable evidence base. One issue that is critical to research on human-robot interaction is cross-platform generalisability. This speaks to the seemingly simple question if similar results are obtained when a different robotic platform is used? Contributions in the section on development and learning for human-robot interaction already successfully focus on cross-platform generalisability. Future studies should continue to pay attention to this issue, as this question is equally important for the development of sophisticated robotic systems as well as studies on the human user (e.g. REF [7]).

The majority of contributions featured in this theme issue focus on the development of or interaction with physically embodied robots. Indeed, besides the impact of artificial intelligence in the digital domain, many of the most pressing societal questions concern the impact that physically embodied artificial agents will have in social spaces (e.g. workplace, education, health care). The focus on physically embodied social robots provides a unique contribution to field, especially in the field of psychology and neuroscience where often digital representation of human or artificial agents are used to study social perception and behaviour. Recent theoretical and empirical accounts suggest that embodiment, as is the case with a robot, is crucial for social cognition and human-robot interaction [2].Most studies on human-robot interaction, investigate real and interactive loops, where a human agent's behaviour triggers reactions in the robot, which trigger behavioural consequences in the human agent, and so on. The focus on physically embodied social robots (or humans) is critical to move towards a true understanding of human social behaviour [19]. When debating about human-robot interaction and robots which are designed to infiltrate the human (social) environment, one cannot forget about the importance of discussion on societal impact of robotics, as well as ethical and moral implications. The vision of today's researchers is to design robots that can assist us in daily chores, in healthcare and elderly care, just to name a few. However, important questions must be addressed, such as how such disruptive technology will change our societies? How will it change our social relationships? What kind of bonding and commitment will we develop in relation to robot companions? How will this change across weeks, months or years of interactions with robot companions or assistants? Can robots substitute "the human touch" in various domains of life? What are the attitudes of humans towards the future with robots, and how can those attitudes evolve with gradual introduction of technology to everyday environments? These and many other questions remain wide open. We hope that the concluding section of this theme issue on "Moral and ethical implications", as well as contributions throughout this issue, can inspire contemplation of such topics, and trigger debate that will lead towards a rigorous scientific approach in addressing the varied and complex issues that require careful consideration in the context of society in the new era – an era where artificial agents take part in our daily living.

Author profiles

Emily S. Cross: Emily is a Professor of Social Robotics and social neuroscientist based at the Institute of Neuroscience and Psychology at the University of Glasgow in Scotland, where she directs the Social Brain in Action Laboratory and serves as PI on the ERC Starting Grant project 'SOCIAL ROBOTS'. Using intensive training procedures, functional neuroimaging, brain stimulation, and research paradigms involving dance, acrobatics and robots, she leads a team who explores how experience-dependent plasticity is manifest in the human brain and behaviour. She is particularly interested in how we learn via observation, how expertise shapes perception, the relationship between embodiment and aesthetics, and social influences on human-robot interaction. Emily received a BA in psychology and dance from Pomona College, an MSc in cognitive psychology from the University of Otago as a Fulbright Fellow, and a PhD in cognitive neuroscience from Dartmouth College. She completed postdoctoral training at the University of Nottingham and the Max Planck Institute for Human Cognitive and Brain Sciences, and was previously an assistant professor at Radboud University Nijmegen and a Professor of Cognitive Neuroscience at Bangor University. She is a 2018 recipient of the Philip Leverhulme Prize in Psychology, and her research has been funded by various national and international organisations, including the NIH (USA), Netherlands Organisation for Scientific Research, Economic and Social Research Council (UK), Ministry of Defence (UK), and the European Research Council.

Ruud Hortensius: Ruud Hortensius is a postdoctoral research fellow in the Institute of Neuroscience and Psychology at the University of Glasgow, Scotland. In his research he investigates the behavioural and brain mechanisms of the rich and diverse social lives of humans. From the communication of emotions to helping behaviour in a social context. To that goal, he uses neuroimaging, brain stimulation, behavioural measures, and virtual reality. Recently, he started to integrate this work with recent developments in social robotics, to study interactions with artificial agents, such as social robots. In this research, he focuses on the flexibility of social cognition, in particular on the impact of long-term interaction with social robots on the neural representation of social cognition. To foster a research community centred around human social cognition in an era of technology, he co-organized two interdisciplinary workshops on the potential of cognitive neuroscience perspectives and techniques for human-robot interaction in 2017 and 2018. Ruud received a BSc in Psychology (cum laude) and a MSc in cognitive neuroscience from Utrecht University in The Netherlands, and a PhD in social and affective neuroscience (cum laude) from Tilburg University in The Netherlands. Previously, he was a postdoctoral researcher at Maastricht University and the University of Cape Town.

Agnieszka Wykowska: Professor Agnieszka Wykowska leads the unit "Social Cognition in Human-Robot Interaction" at the Italian Institute of Technology (Genoa, Italy) and is also affiliated with the Luleå University of Technology, Sweden, as adjunct professor in engineering psychology. In 2016 she was awarded the ERC Starting Grant "Intentional stance for social attunement". Prof. Wykowska studied neuro-cognitive psychology (Ludwig Maximilian University, Munich) and philosophy (Jagiellonian University, Krakow). She obtained PhD in psychology and the German "Habilitation" from the Ludwig Maximilian University. She is an editor-in-chief of the International Journal of Social Robotics, and associate editor of Frontiers in Psychology (section Cognition). She has been serving as program committee member for the conferences: "International Conference on Social Robotics", "Human-Robot Interaction", "Advanced Robotics and its Social Impacts". In 2013, she received an Early Stage Career Prize at the COST meeting "The future concept and reality of social robotics: challenges, perception and applications - role of social robotics in current and future society". Prof. Wykowska has published numerous papers in high-ranking journals across various disciplines. In her research, she examines how humans respond to humanoid robots, and how to make robots' behaviour comprehensible for humans. Apart from contribution to social and cognitive neuroscience, her research contributes to the development of artificial intelligence for social robotics, and the design of robots for societal needs (healthcare, elderly care and daily assistance).

References

1. Samani, H., Saadatian, E., Pang, N., Polydorou, D., Fernando, O. N. N., Nakatsu, R. & Koh, J. T. K. V. 2013 Cultural robotics: the culture of robotics and robotics in culture. *International Journal of Advanced Robotic Systems* **10**, 400.

2. Wykowska, A., Chaminade, T. & Cheng, G. 2016 Embodied artificial agents for understanding human social cognition. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **371**. (doi:10.1098/rstb.2015.0375)

3. Hortensius, R. & Cross, E. S. 2018 From automata to animate beings: the scope and limits of attributing socialness to artificial agents. *Annals of the New York Academy of Sciences* **1426**, 93–110. (doi:10.1111/nyas.13727)

4. In press. Robotics Funding Gets Government Attention in Q1 2017. *httpswww.roboticsbusinessreview.comdownloadrobotics-funding-gets-government-attention-q*.

5. Broadbent, E. 2017 Interactions With Robots: The Truths We Reveal About Ourselves. *Annu. Rev. Psychol.* **68**, 627–652. (doi:10.1146/annurev-psych-010416-043958)

6. Chaminade, T. & Cheng, G. 2009 Social cognitive neuroscience and humanoid robotics. *Journal of Physiology - Paris* **103**, 286–295. (doi:10.1016/j.jphysparis.2009.08.011)

7. Hortensius, R., Hekele, F. & Cross, E. S. In press. The perception of emotion in artificial agents. *IEEE Trans. Cogn. Dev. Syst.*, 1–1. (doi:10.1109/TCDS.2018.2826921)

8. Wiese, E., Metta, G. & Wykowska, A. 2017 Robots As Intentional Agents: Using Neuroscientific Methods to Make Robots Appear More Social. *Front. Psychol.* **8**, 1663. (doi:10.3389/fpsyg.2017.01663)

9. Metta, G. et al. 2010 The iCub humanoid robot: An open-systems platform for research in cognitive development. *Neural Networks* **23**, 1125–1134.

10. Cangelosi, A. & Schlesinger, M. 2018 From Babies to Robots: The Contribution of Developmental Robotics to Developmental Psychology. *Child Dev Perspect* **12**, 183–188. (doi:10.1111/cdep.12282)

11. Meltzoff, A. N. 2007 'Like me': a foundation for social cognition. *Developmental Science* **10**, 126–134. (doi:10.1111/j.1467-7687.2007.00574.x)

12. Ogawa, K., Nishio, S., Koda, K., Balistreri, G., Watanabe, T. & Ishiguro, H. 2011 Exploring the natural reaction of young and aged person with telenoid in a real world. *JACIII* **15**, 592–597.

13. Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. & Munafò, M. R. 2013 Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* **14**, 365.

14. Collaboration, O. S. 2015 Estimating the reproducibility of psychological science. *Science* **349**, aac4716.

15. Collberg, C., Proebsting, T. & Warren, A. M. 2015 Repeatability and benefaction in computer systems research. *University of Arizona TR 14* **4**.

16. Hutson, M. 2018 Artificial intelligence faces reproducibility crisis. *Science* **359**, 725–726. (doi:10.1126/science.359.6377.725)

17. Nelson, L. D., Simmons, J. & Simonsohn, U. 2018 Psychology's Renaissance. *Annu. Rev. Psychol.* **69**, 511–534. (doi:10.1146/annurev-psych-122216-011836)

18. Munafò, M. R. et al. 2017 A manifesto for reproducible science. *Nature Publishing Group* **1**, 0021. (doi:10.1038/s41562-016-0021)

19. Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T. & Vogeley, K. 2013 Toward a second-person neuroscience. *Behav Brain Sci* **36**, 393–414. (doi:10.1017/S0140525X12000660)