From the Human Brain to Robots and Back Again: Exploring Mutually Beneficial Exchange between Social Cognition, Neuroscience and Robotics

Emily S. Cross

Abstract— **A core challenge for social cognition research concerns understanding how we perceive and interact with others in a social world. With the growing presence of artificial agents in society, including humanoid robots and computer-generated avatars, this challenge promises to become even more complex, as researchers work to determine the extent to which neurocognitive mechanisms evolved to support interactions with other people are co-opted during interactions with artificial agents. While a growing number of researchers working within the domain of social robotics are building a case for the utility of social cognition and social neuroscience research, it is becoming increasingly evident that those working within social cognition and neuroscience domains stand to benefit just as much through collaborations with social roboticists. This paper examines how social robots can be used to help researchers working within social cognition and human neuroscience domains examine fundamental questions of the discipline, with mutual benefits feeding back to those interested in developing robots for socially interactive public settings.**

I. INTRODUCTION

"The goal of creating robots that can engage with us as full-fledged partners is as challenging and deep as artificial intelligence's original goal, because it requires scientists of the natural and the artificial to deeply understand human intelligence, behavior and nature across multiple dimensions (cognitive, affective, physical and social)." – Dr. Cynthia Breazeal, director, Personal Robots Group at the MIT MediaLab

Social interactions form the foundation of human society, as our abilities to perceive, respond to, and coordinate behaviour with others are necessary for us to survive and thrive in a social world. A major challenge for psychologists and those working in related disciplines is to characterize *how* we understand and coordinate our actions with others to achieve mutual goals. While breakthroughs in social cognition and social neuroscience research are helping to elucidate the mechanisms and consequences of social perception and interaction [1], this challenge is poised to intensify with the growing presence of artificial agents in society. Nearly a decade ago, Microsoft founder Bill Gates prophesized a revolution in the robotics industry that would see staggering leaps in the progress and sophistication of robots, and predicted "a robot in every home" in the near future [2]. While the ubiquity of home robots has yet to be realised, in 2012-13 alone, more than 270 robotics-related start-up companies emerged globally, with a growing percentage focusing on the development of companion robots for the home or assistance robots to serve in complex, human-interactive contexts, like schools, hospitals and care homes [3].

As interest in and development of such social robots continues apace [4-8], our understanding of the extent to which we perceive and interact with these machines as social agents, as evidenced by brain and behavioural responses, remains rudimentary. Investigating how neurocognitive systems that have evolved to support social interactions with other humans are co-opted to support interactions with social robots will have important implications not only for the design and programming of socially interactive robotic agents, but equally importantly, for our understanding of the flexibility of neurocognitive processes supporting social behavior in general.

To better understand how people might respond to artificial agents in social settings, it is imperative to understand in more detail the hardware supporting social behavior: the human brain. The tools and techniques offered by social neuroscience, an emerging field that combines aspects of social psychology, social cognition, and human neuroscience, offer a particularly promising avenue for exploring the human side of human-robot interactions, as evidenced by a growing body of empirical work in this domain [4-6]. The aim of this paper is strengthen the foundation for reciprocal exchange between the domains of social cognition/social neuroscience and social robotics. To achieve this, this paper is organized in such a manner that core principles of social cognition research are introduced and then linked to relevant findings from social neuroscience concerning the neurobiology of social perception and interaction. The next section highlights research investigating how humans perceive and respond to artificial agents in social settings, with an emphasis on findings that challenge a foundational principle of social cognition, the 'like me' hypothesis. The final section articulates several benefits from developing further exchange between scientists working with the human brain and artificial intelligence.

As a researcher working squarely within social cognition and social neuroscience domains, it is important to clarify that my experience with social robotics to date has focused on the utility of artificial agents to deepen our understanding of the social nature of the human brain and behavior. However, as work by my team and a growing number of other human neuroscience laboratories demonstrates (e.g., [9-20]), continued crosstalk and collaboration between researchers working from human-focused and robotics-

*Research supported by the Economic and Social Research Council, UK (ES/K001892/1), the Ministry of Defense - Defense Science and Technology Laboratory (DSTLX-1000008317725), the Netherlands Organisation for Scientific Research (Veni Award #451-11-002) and a FP7 Marie Curie Career Integration Grant (PCIG11-GA-2012-322256).

E.C. is currently at the Social Brain in Action Laboratory at the School of Psychology at Bangor University, Bangor, LL57 2AS, Wales (e-mail: e.cross@bangor.ac.uk), and was previously at the Department of Social and Cultural Psychology, Behavioural Science Institute, and Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, the Netherlands.

focused perspectives should result in the dual benefits of improved design and uptake of social robotics, as well as advanced understanding of the neurobiological hardware supporting social exchange.

II. Linking Social Cognition with the Social Brain

The domain of social cognition provides a useful point of departure for exploring how we perceive and interact with other agents, whether biological or artificial. A dominant theoretical stance in social cognition stipulates that understanding the basic similarity between self and others forms the foundation of social cognition, and that humans have developed to seek out self-other equivalence in others [21, 22]. This account, known as the 'like me' hypothesis, further proposes that actions performed by oneself and another are represented in common cognitive codes [21]. The discovery of mirror neurons in premotor and parietal cortices of the monkey brain bolstered this idea by providing evidence at the cellular level for overlapping representations of actions performed by self and others [23, 24].

Behavioral and neuroimaging studies performed with human subjects provide additional support for the 'like me' hypothesis. This work demonstrates evidence of behavioral facilitation [17, 25] and increased engagement of mirror system brain regions when individuals observe familiar actions or interact with agents similar to themselves [19, 26-29]. While the mirror system might support the sharing of experience between actor and observer, successful interaction with others also entails taking an interaction partner's perspective. The temporoparietal junction (TPJ) is implicated in this process [30, 31], and neuroimaging findings support the 'like-me' nature of TPJ engagement when interacting with socially similar others [15, 32]. Together, the mirror system and TPJ are core components of a broader network of cortical regions collectively called the 'social brain' [1, 33] (see Fig. 1). A growing body of social neuroscience research continues to explore the extent to which perception of 'like me'-ness forms the bedrock of social cognition [4, 13, 26, 34-37]. As the next section illustrates, the domain of social robotics is helping to test key questions concerning minimum standards for social perception and engagement.

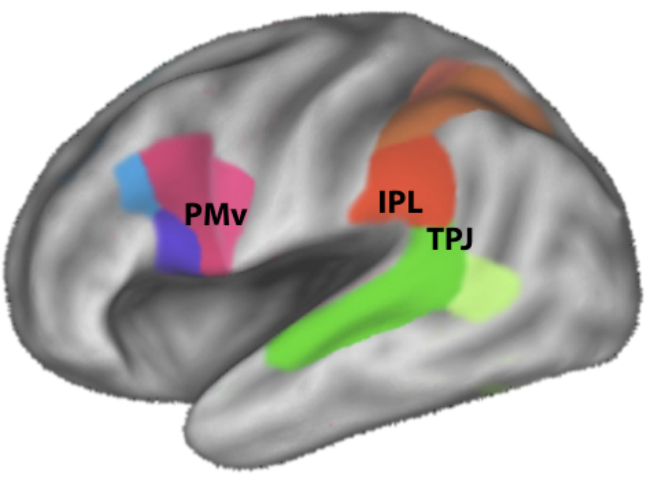

Fig. 1. The Social Brain. This inflated brain image illustrates the core brain regions implicated in social perception and cognition: PMv = ventral premotor cortex; IPL = inferior parietal lobule; TPJ = temporoparietal junction. *(note that these brain regions are distributed bilaterally, but are depicted here in the left hemisphere only)*

III. From Social Neuroscience to Social Robotics

Over the past decade, individuals working to develop socially interactive robots have taken an increased interest in the social cognition and social neuroscience research reviewed above [5, 34, 38-40]. An on-going goal for robotics designers has been to maximize the similarity of artificial agents to humans, in terms of appearance and movement (while perhaps attempting to circumnavigate the uncanny valley[1]), in an aim to make particular artificial agents as 'like me' as possible [6]. However, recent data raise a challenge to the notion that the more 'like me' another agent looks or moves, the better a person can understand or interact with it [11].

Using functional magnetic resonance imaging (fMRI), my team has demonstrated that core nodes of the social brain responded more strongly to movements performed by a human or robotic agent that featured more stereotypically robotic kinematics compared to more smooth, human-like kinematics [11] (see Fig. 2). A remarkably similar pattern of brain activity also emerged within the brains of 4-month-old infants (this time, scanned using functional near infrared spectroscopy, a baby-friendly neuroimaging approach) when they watched the identical human and robot videos [41]. These data suggest that not only do brain regions associated with social perception respond strongly when observing non-human motion kinematics, but also that this flexibility in information process is present from birth. Such findings open a range of new questions concerning the role of experience with biological and artificial agents in changing perception, and what is innate in terms of brain mechanisms supporting perception of other agents.

New data by my team also demonstrates that the form and motion of an artificial agent might in fact be less crucial for engaging brain regions associated with social cognition than the extent to which a human interaction partner *believes* the agent to have human origins [15, 42]. For example, we recently used an automatic imitation task [43] to examine the extent to which participants imitated the hand actions of human-like or robotic-like virtual agents based on whether the actions were believed to have originated from human motion capture technology or purely computer-driven animation techniques [15] (Fig. 3). Behaviorally, we found that participants imitated the virtual hand actions equivalently when either cue to humanness was present (either when the hand looked human-like or when

[1] The 'Uncanny Valley' is a frequently cited theory first proposed in the 1970s that stipulates that people report increasing feelings of fondness for an artificial agent the more human-like it looks or acts, but only up to a certain point. Once an artificial agent appears too similar to a human, feelings of fondness quickly turn to feelings of revulsion [49]. As evocative and widely-cited as this theory is, lively debate continues concerning whether the Uncanny Valley theory is supported by empirical findings [50-52].

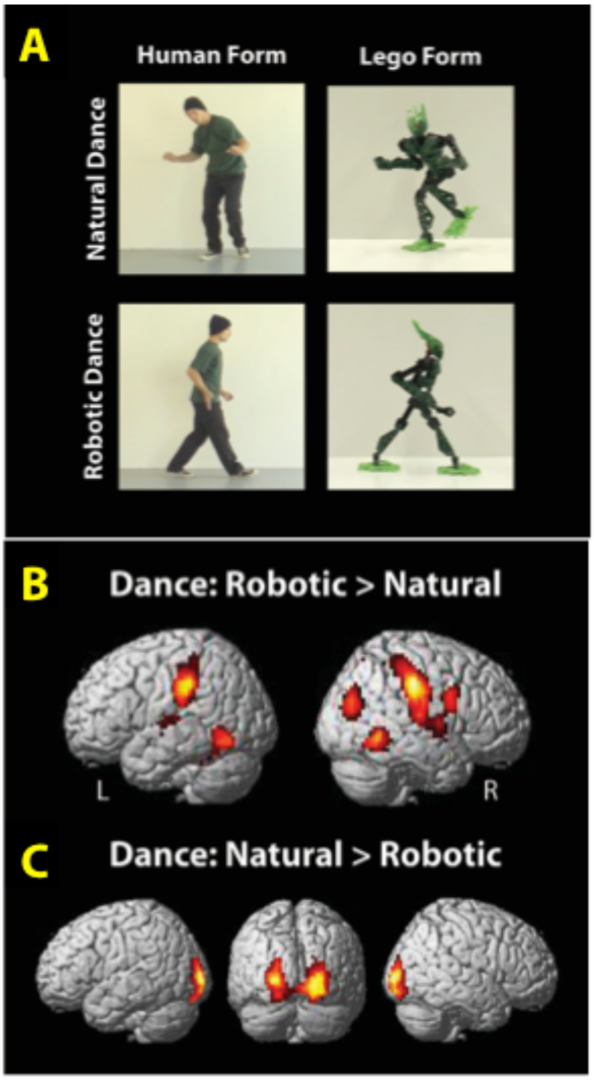

Fig. 2. Experimental design, stimuli and results from study investigating the impact of robotic form and motion on engagement of social brain regions. Panel A depicts the research design, which examined both human and roboitc form and motion cues using videos of a person or Lego robot dancing. Panel B illustrates widespread engagement of social brain regions when participants watched the human or robot agent move in a rigid, mechanistic way. Panel C shows that when comparing more human-like to more robot-like motion, only early visual regions of the brain are engaged, contrary to what the 'like me' hypothesis might predict. Figure adapted from [11].

participants believed it originated from motion capture technology). In terms of brain activity, we saw engagement of TPJ only when both cues to human animacy were present. Taken together, the findings reported by my team [11, 15, 41, 42] and others [14, 20, 44] are thus starting to call into question the notion that core social nodes of the human brain have evolved to respond most to agents that look, move, and are generally believed to be "like me" [21, 22]. As roboticists interested in developing artificial agents for public social settings (such as schools, assisted living facilities, and the workplace) begin to draw more upon social neuroscience to improve human-robot interactions [5, 8], it will become increasingly important for close collaboration and exchange between disciplines to ensure the latest findings from basic social cognition and social neuroscience research can effectively inform the design and introduction of social robots.

IV. Benefits of Collaborative Exchange Between Social Cognition, Neuroscience and Robotics

To date, a number of researchers working within robotics and neurocognitive domains have made a strong case for the utility of collaborative exchange between these disciplines [4-8, 13, 20, 35]. Perhaps not surprisingly given the inherently interdisciplinary nature of robotics (drawing upon computing, engineering, science and mathematics), social roboticists appear to be better versed at borrowing from social neurocognitive research to advance their aims than the other way around. To give a specific example, in a recent technical correspondence in *IEEE Transactions on Human-Machine Systems*, Wiltshire and Fiore provide a detailed overview of three key areas of social, cognitive, and affective neuroscience (SCAN) research that hold particularly rich promise for human-machine system research: emotion, theory of mind, and joint action [8]. Drawing on SCAN research in these three areas, the authors present a compelling case for using research examining human behavioral and brain responses (when perceiving and interacting mostly with other humans) to enrich research and development in the human-machine interface domains of learning and training, human-robot interactions, and enhancing team performance. The focus of this paper, and indeed, related works in the robotics domain [4-6, 36, 38, 39, 45], is on harnessing insights and research methods from social neurocognitive approaches to improve the development of social (and sociable) robots.

The flip side of this interdisciplinary exchange, however, concerns the untapped utility of artificial agents to facilitate the research efforts of those working within social cognition and social neuroscience. As reviewed in the previous section, a number of laboratories are examining the extent to which self-other equivalencies form the basis for social cognition (as proposed by the 'like me' hypothesis [21, 22]). Over the past decade, pioneering work by Thierry Chaminade [4, 5, 10] and Ayse Sagin [18, 46, 47] has drawn upon social robotics and computer-generated artificial agents in a particularly innovative manner to examine core questions for social neurocognitive researchers, including attributions of agency, perception of biological motion, how strong people's anthropomorphic bias truly is, and the extent to which notions of the uncanny valley are validated by empirical evidence. More recently, my team and others have been using artificial agents to explore the biological tuning of brain regions implicated in social cognition [11, 14, 15, 42, 44]. Together, findings from this work are starting to raise challenges to foundational theories of social cognition, which should be seen as an exciting opportunity for further cross-disciplinary experimentation using artificial agents.

From a social neurocognitive perspective, the time is also ripe to take a more detailed look into the role played by the social brain when interacting with dissimilar others, such as artificial agents. Important questions remain concerning the extent to which brain regions supporting interaction with

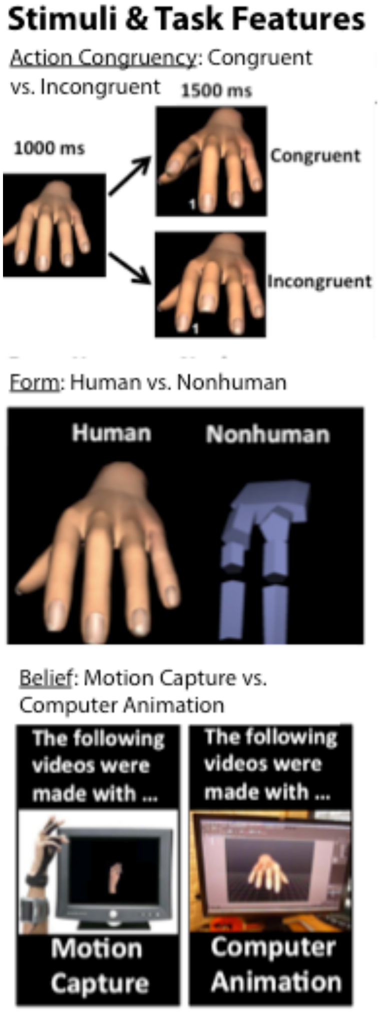

Fig 3. Experimental conditions for interactive imitation task that human participants performed whilst undergoing functional neuroimaging. Participants interacted with a hand that could make finger movements that were either congruent or incongruent to the participants' own movements; the hand could either have a human or a non-human form; and finally, participants were told that they were either interacting with hands that were generated via human motion capture technology or computer animation. Figure adapted from [15].

humans overlap or diverge with those supporting interaction with robots, whether the extent of overlap determines the perceived socialness of artificial agents, and the influence of experience. Answers to these questions are necessary to establish the extent to which the social brain can code robots as 'like me' (and the extent to which this is necessary or desirable), as a way to enhance and guide future robotic development. Robotics experts recognize that understanding social cognition and certain aspects of human neuroscience is essential for building interactive robots [45, 48]. Ongoing investigations using the tools and approaches of social neuroscience will serve to further inform this field by establishing how experience with humanoid robots shapes perception about their social characteristics, and the extent to which a robot must be perceived as 'like me' to elicit truly social responses in a human interaction partner.

References

[1] Frith, C.D.: 'The social brain?', Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 2007, 362, (1480), pp. 671-678

[2] Gates, B.: 'A robot in every home', Scientific American, 2007, 296, (1), pp. 58-65

[3] Tobe, F.: 'A glimpse at our robotic future: 235 start-ups reviewed', in Editor (Ed.)^(Eds.): 'Book A glimpse at our robotic future: 235 start-ups reviewed' (2013, edn.), pp. 191-198

[4] Chaminade, T.: 'Acquiring and probing self-other equivalencies Using artificial agents to study social cognition', in Editor (Ed.)^(Eds.): 'Book Acquiring and probing self-other equivalencies Using artificial agents to study social cognition' (2006, edn.), pp. 769-774

[5] Chaminade, T., and Cheng, G.: 'Social cognitive neuroscience and humanoid robotics', J Physiol Paris, 2009, 103, (3-5), pp. 286-295

[6] Coradeschi, S., Ishiguro, H., Asada, M., Shapiro, S.C., Thielscher, M., Breazeal, C., and Mataric, M.: 'Human-Inspired Robots', IEEE Intelligent Systems, 2006, 21, pp. 74-85

[7] Hanson, D., Mazzei, D., Garver, C., De Rossi, D., and Stevenson, M.: 'Realistic humanlike robots for treatment of ASD, social training and research appeal to youths with ASD, cause physiological arousal, and increase human-to-human social engagement', in Editor (Ed.)^(Eds.): 'Book Realistic humanlike robots for treatment of ASD, social training and research appeal to youths with ASD, cause physiological arousal, and increase human-to-human social engagement' (2012, edn.), pp.

[8] Wiltshire, T.J., and Fiore, S.M.: 'Social Cognitive and Affective Neuroscience in Human–Machine Systems: A Roadmap for Improving Training, Human–Robot Interaction, and Team Performance', Human-Machine Systems, IEEE Transactions on, 2014, 44, (6), pp. 779-787

[9] Bird, G., Leighton, J., Press, C., and Heyes, C.: 'Intact automatic imitation of human and robot actions in autism spectrum disorders', Proceedings. Biological sciences / The Royal Society, 2007, 274, (1628), pp. 3027-3031

[10] Chaminade, T., Franklin, D.W., Oztop, E., and Cheng, G.: 'Motor interference between humans and humanoid robots: Effect of biological and artifical motion', Proceedings of the 2005 4th IEEE International Conference on Development and Learning, 2005, pp. 96-101

[11] Cross, E.S., Liepelt, R., de, C.H.A.F., Parkinson, J., Ramsey, R., Stadler, W., and Prinz, W.: 'Robotic movement preferentially engages the action observation network', Hum Brain Mapp, 2012, 33, (9), pp. 2238-2254

[12] Cross, E.S., Stadler, W., Parkinson, J., Schutz-Bosbach, S., and Prinz, W.: 'The influence of visual training on predicting complex action sequences', Hum Brain Mapp, 2013, 34, (2), pp. 467-486

[13] Epley, N., Waytz, A., and Cacioppo, J.T.: 'On seeing human: a three-factor theory of anthropomorphism', Psychol Rev, 2007, 114, (4), pp. 864-886

[14] Gazzola, V., Rizzolatti, G., Wicker, B., and Keysers, C.: 'The anthropomorphic brain: the mirror neuron system responds to human and robotic actions', Neuroimage, 2007, 35, (4), pp. 1674-1684

[15] Klapper, A., Ramsey, R., Wigboldus, D., and Cross, E.S.: 'The control of automatic imitation based on bottom-up and top-down cues to animacy: insights from brain and behavior', J Cogn Neurosci, 2014, 26, (11), pp. 2503-2513

[16] Press, C.: 'Action observation and robotic agents: learning and anthropomorphism', Neuroscience and biobehavioral reviews, 2011, 35, (6), pp. 1410-1418

[17] Press, C., Bird, G., Flach, R., and Heyes, C.: 'Robotic movement elicits automatic imitation', Brain Res Cogn Brain Res, 2005, 25, (3), pp. 632-640

[18] Saygin, A.P., and Stadler, W.: 'The role of appearance and motion in action prediction', Psychological research, 2012, 76, (4), pp. 388-394

[19] Shimada, S.: 'Deactivation in the sensorimotor area during observation of a human agent performing robotic actions', Brain Cogn, 2010, 72, (3), pp. 394-399

[20] Wang, Y., and Quadflieg, S.: 'In our own image? Emotional and neural processing differences when observing human-human vs human-robot interactions', Social cognitive and affective neuroscience, 2015

[21] Meltzoff, A.: ''Like me': a foundation for social cognition', Developmental science, 2007, 10, (1), pp. 126-134
[22] Meltzoff, A., and Prinz, W.: 'The imitative mind: Development, evolution, and brain bases' (Cambridge University Press, 2003. 2003)
[23] Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G.: 'Action recognition in the premotor cortex', Brain, 1996, 119 (Pt 2), pp. 593-609
[24] Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L.: 'Premotor cortex and the recognition of motor actions', Brain Res Cogn Brain Res, 1996, 3, (2), pp. 131-141
[25] Catmur, C., Walsh, V., and Heyes, C.: 'Sensorimotor learning configures the human mirror system', Curr Biol, 2007, 17, (17), pp. 1527-1531
[26] Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., Porro, C.A., and Rizzolatti, G.: 'Neural circuits involved in the recognition of actions performed by nonconspecifics: an FMRI study', J Cogn Neurosci, 2004, 16, (1), pp. 114-126
[27] Cross, E.S., Hamilton, A.F., and Grafton, S.T.: 'Building a motor simulation de novo: Observation of dance by dancers', Neuroimage, 2006, 31, (3), pp. 1257-1267
[28] Cross, E.S., Kraemer, D.J., Hamilton, A.F., Kelley, W.M., and Grafton, S.T.: 'Sensitivity of the action observation network to physical and observational learning', Cereb Cortex, 2009, 19, (2), pp. 315-326
[29] Tai, Y.F., Scherfler, C., Brooks, D.J., Sawamoto, N., and Castiello, U.: 'The human premotor cortex is 'mirror' only for biological actions', Curr Biol, 2004, 14, (2), pp. 117-120
[30] Apperly, I.A., Samson, D., Chiavarino, C., and Humphreys, G.W.: 'Frontal and temporo-parietal lobe contributions to theory of mind: neuropsychological evidence from a false-belief task with reduced language and executive demands', J Cogn Neurosci, 2004, 16, (10), pp. 1773-1784
[31] Saxe, R., and Kanwisher, N.: 'People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind"', Neuroimage, 2003, 19, (4), pp. 1835-1842
[32] Takahashi, H., Terada, K., Morita, T., Suzuki, S., Haji, T., Kozima, H., Yoshikawa, M., Matsumoto, Y., Omori, T., Asada, M., and Naito, E.: 'Different impressions of other agents obtained through social interaction uniquely modulate dorsal and ventral pathway activities in the social human brain', Cortex; a journal devoted to the study of the nervous system and behavior, 2014, 58, pp. 289-300
[33] Pelphrey, K.A., and Carter, E.J.: 'Brain mechanisms for social perception: lessons from autism and typical development', Annals of the New York Academy of Sciences, 2008, 1145, pp. 283-299
[34] Asada, M.: 'Cognitive developmental robotics as a new paradigm for the design of humanoid robots', Robotics and Autonomous Systems, 2001, 37, pp. 185-193
[35] Hofree, G., Ruvolo, P., Bartlett, M.S., and Winkielman, P.: 'Bridging the mechanical and the human mind: spontaneous mimicry of a physically present android', PloS one, 2014, 9, (7), pp. e99934
[36] Kiesler, S., Powers, A., Fussell, S.R., and Torrey, C.: 'Anthropomorphic interactions with a robot and a robot-like agent', Social Cognition, 2008, 26, (2), pp. 169-181
[37] Liepelt, R., and Brass, M.: 'Top-down modulation of motor priming by belief about animacy', Experimental psychology, 2010, 57, (3), pp. 221-227
[38] Breazeal, C.: 'Sociable robots', Journal of the Robotics Society of Japan, 2007, 24, (5), pp. 591-593
[39] Ishiguro, H.: 'Studies on very humanlike robots', in Editor (Ed.)^(Eds.): 'Book Studies on very humanlike robots' (2013, edn.), pp.
[40] Asada, M.: 'Development of artificial empathy', Neuroscience research, 2014
[41] Grossmann, T., Cross, E.S., Ticini, L.F., and Daum, M.M.: 'Action observation in the infant brain: the role of body form and motion', Social neuroscience, 2013, 8, (1), pp. 22-30
[42] Cross, E.S., Ramsey, R., Liepelt, R., Prinz, W., and Hamilton, A.F.: 'The shaping of social perception by stimulus and knowledge cues to human animacy', under review
[43] Brass, M., Bekkering, H., and Prinz, W.: 'Movement observation affects movement execution in a simple response task', Acta Psychol (Amst), 2001, 106, (1-2), pp. 3-22
[44] Ramsey, R., and Hamilton, A.F.d.C.: 'Triangles have goals too: understanding action representation in left aIPS', Neuropsychologia, 2010, 48, (9), pp. 2773-2776
[45] Ishihara, H., and Asada, M.: 'Five key characteristics for intimate human-robot interaction: Developmen of uppoer torso for child robot 'Affetto'', Adv Robotics, 2014, in press
[46] Saygin, A.P., Chaminade, T., Ishiguro, H., Driver, J., and Frith, C.: 'The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions', Social cognitive and affective neuroscience, 2012, 7, (4), pp. 413-422
[47] Urgen, B.A., Plank, M., Ishiguro, H., Poizner, H., and Saygin, A.P.: 'EEG theta and Mu oscillations during perception of human and robot actions', Frontiers in neurorobotics, 2013, 7, pp. 19
[48] Lakatos, G., Gacsi, M., Konok, V., Bruder, I., Bereczky, B., Korondi, P., and Miklosi, A.: 'Emotion attribution to a non-humanoid robot in different social situations', PloS one, 2014, 9, (12), pp. e114207
[49] Mori, M. On the Uncanny Valley. Proceedings of the Humanoids-2005 Workshop: Views of the Uncanny Valley, 2005, Tuskuba, Japan
[50] Ceetham, M., Suter, P. and Janke, L.: 'Perceptual disiniation diffuclty and familiarity in the Uncanny Valley: More like a "Happy Valley", Front Psycho, 2015, pp. 1219
[51] Katsyri, J., Forger, K., Makarainen, M. and Takala, T.: 'A review of empiricalevidence on different uncanny valley hypotheses: Support for perceptual mismatch as one road to the valley of eeriness', Front Psychol, 2015, 6, pp.390
[52] Piwek, L., McKay, L.S. and Pollick, F. E.:'Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion', Cognition, 2014, 130(3), pp. 271-277