frontiers in Robotics and AI

OPINION
published: 19 February 2021
doi: 10.3389/frobt.2021.633514

Can You Activate Me? From Robots to Human Brain

F. Manzi[1,2,3]*, *C. Di Dio*[1,2,3], *D. Di Lernia*[2,3], *D. Rossignoli*[2,4,5], *M. A. Maggioni*[2,4,5], *D. Massaro*[1,2,3], *A. Marchetti*[1,2,3] and *G. Riva*[2,3,6]

[1]Research Unit on Theory of Mind, Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy, [2]Humane Technology Lab, Università Cattolica del Sacro Cuore, Milan, Italy, [3]Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy, [4]DISEIS, Department of International Economics, Institutions and Development, Universitá Cattolica del Sacro Cuore, Milan, Italy, [5]CSCC, Cognitive Science and Communication research Center, Universitá Cattolica del Sacro Cuore, Milan, Italy, [6]Applied Technology for NeuroPsychology Laboratory, Istituto Auxologico Italiano, Milan, Italy

Keywords: human-robot interaction, social brain, long-term interaction, fMRI, EEG

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 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 cortex 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, Stern, 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

OPEN ACCESS

Edited by:
Ginevra Castellano, Uppsala University, Sweden

Reviewed by:
Maria Alessandra Umilta, University of Parma, Italy

***Correspondence:**
F. Manzi
federico.manzi@unicatt.it

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *25 November 2020*
Accepted: *15 January 2021*
Published: *19 February 2021*

Citation:
Manzi F, Di Dio C, Di Lernia D, Rossignoli D, Maggioni MA, Massaro D, Marchetti A and Riva G (2021) Can You Activate Me? From Robots to Human Brain. Front. Robot. AI 8:633514. doi: 10.3389/frobt.2021.633514

FIGURE 1 | Robots can activate the human brain.

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 heterogenous 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 and colleagues (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 anthropomorphize 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)

AUTHOR CONTRIBUTIONS

FM conceived the idea. FM, DDL and DR selected the articles. FM, CD, and GR finalised the idea. All authors contributed to the article.

FUNDING

This work has been supported by Universitá Cattolica del Sacro Cuore, Milan, (D3.2 - 2018 - Human-Robot Confluence project).

REFERENCES

Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., and Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proc. Natl. Acad. Sci. U.S.A.* 100 (9), 5497–5502. doi:10.1073/pnas.0935845100

Casile, A., Dayan, E., Caggiano, V., Hendler, T., Flash, T., and Giese, M. A. (2009). Neuronal encoding of human kinematic invariants during action observation. *Cerebr. Cortex* 20, 1647–1655. doi:10.1093/cercor/bhp229

Chaminade, T., Zecca, M., Blakemore, S. J., Takanishi, A., Frith, C. D., Micera, S., et al. (2010). Brain response to a humanoid robot in areas implicated in the perception of human emotional gestures. *PloS One*, 5 (7), e11577. doi:10.1371/journal.pone.0011577

Cross, E. S., Hortensius, R., and Wykowska, A. (2019). From social brains to social robots: applying neurocognitive insights to human–robot interaction. *Phil. Trans. Biol. Sci.* 374 (1771), 20180024. doi:10.1098/rstb.2018.0024

Dayan, E., Casile, A., Levit-Binnun, N., Giese, M. A., Hendler, T., and Flash, T. (2007). Neural representations of kinematic laws of motion: evidence for action-perception coupling. *Proc. Natl. Acad. Sci. U.S.A.* 104 (51), 20582–20587. doi:10.1073/pnas.0710033104

Di Cesare, G., Valente, G., Di Dio, C., Ruffaldi, E., Bergamasco, M., Goebel, R., et al. (2016). Vitality forms processing in the insula during action observation: a multivoxel pattern analysis. *Front. Hum. Neurosci.*, 10, 267. doi:10.3389/fnhum.2016.00267

Di Cesare, G., Vannucci, F., Rea, F., Sciutti, A., and Sandini, G. (2020). How attitudes generated by humanoid robots shape human brain activity. *Sci. Rep.*, 10 (1), 16928. doi:10.1038/s41598-020-73728-3

Di Dio, C., Di Cesare, G., Higuchi, S., Roberts, N., Vogt, S., and 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. doi:10.1016/j.neuroimage.2012.09.026

Gazzola, V., Rizzolatti, G., Wicker, B., and Keysers, C. (2007). The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *Neuroimage* 35 (4), 1674–1684. doi:10.1016/j.neuroimage.2007.02.003

Guo, F., Li, M., Qu, Q., and Duffy, V. G. (2019). The effect of a humanoid robot's emotional behaviors on users' emotional responses: evidence from pupillometry and electroencephalography measures. *Int. J. Hum. Comput. Interact.*, 35 (20), 1947–1959. doi:10.1080/10447318.2019.1587938

Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annu. Rev. Psychol.* 60, 653–670. doi:10.1146/annurev.psych.60.110707.163604

Jung, C. E., Strother, L., Feil-Seifer, D. J., and Hutsler, J. J. (2016). Atypical asymmetry for processing human and robot faces in autism revealed by fNIRS. *PloS One* 11 (7), e0158804. doi:10.1371/journal.pone.0158804

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., and Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PloS One*, 3 (7), e2597. doi:10.1371/journal.pone.0002597

Marchetti, A., Manzi, F., Itakura, S., and Massaro, D. (2018). Theory of mind and humanoid robots from a lifespan perspective. *Zeitschrift Für Psychologie* 226 (2), 98–109. doi:10.1027/2151-2604/a000326

Matsuda, G., Hiraki, K., Ishiguro, H., Matsuda, G., Hiraki, K., and Ishiguro, H. (2016). EEG-based mu rhythm suppression to measure the effects of appearance and motion on perceived human likeness of a robot. *J. of Human-Robot Interction* 5 (1), 68–81. doi:10.5898/JHRI.5.1.Matsuda

Paetzel, M., Perugia, G., and 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, 24-26 March 2020, Cambridge, UK, 73–82. doi:10.1145/3319502.3374786

Riva, G., Wiederhold, B. K., Chirico, A., Di Lernia, D., Mantovani, F., and Gaggioli, A. (2018). Brain and virtual reality: what do they have in common and how to exploit their potential. *Annu. Rev. Cyberther. Telemed.*, 16, 3–7.

Riva, G., Wiederhold, B. K., Di Lernia, D., Chirico, A., Riva, E. F. M., Mantovani, F., et al. (2019). Virtual reality meets artificial intelligence: the emergence of advanced digital therapeutics and digital biomarkers. *Annu. Rev. Cyberther. Telemed.*, 17, 3–7.

Rosenthal-von der Pütten, A. M., Schulte, F. P., Eimler, S. C., Sobieraj, S., Hoffmann, L., Maderwald, S., et al. (2014). Investigations on empathy towards humans and robots using fMRI. *Comput. Hum. Behav.* 33, 201–212. doi:10.1016/j.chb.2014.01.004

Rossi, A., Dautenhahn, K., Koay, K. L., Walters, M. L., and Holthaus, P. (2020). "Evaluating people's perceptions of trust in a robot in a repeated interactions study," in *Social robotics*. Editors A. R. Wagner, D. Feil-Seifer, K. S. Haring, S. Rossi, T. Williams, H. He, et al. (New York, NY: Springer International Publishing), 453–465. doi:10.1007/978-3-030-62056-1_38

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., and Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Soc. Cognit. Affect Neurosci.* 7 (4), 413–422. doi:10.1093/scan/nsr025

Stern, D. N. (2010). *Forms of vitality*. Oxford, England: Oxford University Press.

Stern, D. N. (1985). *The interpersonal world of the infant*. New York, NY: Basic Books.

Suzuki, Y., Galli, L., Ikeda, A., Itakura, S., and Kitazaki, M. (2015). Measuring empathy for human and robot hand pain using electroencephalography. *Sci. Rep.* 5 (1), 15924. doi:10.1038/srep15924

Urgen, B. A., Plank, M., Ishiguro, H., Poizner, H., and Saygin, A. P. (2013). EEG theta and Mu oscillations during perception of human and robot actions. *Front. Neurorob.* 7, 19. doi:10.3389/fnbot.2013.00019

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., and Rizzolatti, G. (2003). Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. *Neuron*, 40 (3), 655–664. doi:10.1016/s0896-6273(03)00679-2

Conflict of Interest: 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.

Copyright © 2021 Manzi, Di Dio, Di Lernia, Rossignoli, Maggioni, Massaro, Marchetti and Riva. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.