Using fNIRS to assess the effects of robot errors on social cognition during naturalistic human-robot interaction

Yigit Topoglu[1], Shawn Joshi[1], Ewart de Visser[1,2], Frank Krueger[3], Hasan Ayaz[1,4,5,6,7]

[1] School of Biomedical Engineering, Science and Health Systems, Drexel University, Philadelphia, PA USA

[2] Warfighter Effectiveness Research Center, U.S. Air Force Academy, Colorado Springs, CO USA

[3] School of Systems Biology, George Mason University, Fairfax, VA USA

[4] Drexel Solutions Institute, Drexel University, Philadelphia, PA USA

[5] Department of Psychology, Drexel University, Philadelphia, PA USA

[6] Department of Family and Community Health, University of Pennsylvania, Philadelphia PA USA

[7] Center for Injury Research and Prevention, Children's Hospital of Philadelphia, Philadelphia, PA USA

As autonomous agents get more integrated into our everyday lives, the need for sociable robots that can engage interactively with humans is increasing. For more sociable human-robot interactions, autonomous agents must obtain the appropriate social mechanisms that enable them to be more natural and intuitive with humans[1]. However, the social repertoire in today's robots remains limited and has yet to deliver the expectations[2]. Therefore, integrating robots in human-robot interactions (HRI) as a social entity remains one of the biggest challenges in state-of-art HRI research[3]. For over two decades, HRI researchers have been analyzing human behavior while interacting with robots to find the factors that optimize the user-robot interaction. The most commonly employed methods in HRI studies are subjective self-reported measures such as questionnaires[4]. While these measures are easier to implement, they cannot capture the user's behavior in the moment of interaction and can be dampened through the user's social inhibition. Furthermore, even though neuroimaging methods (e.g. fMRI) have been previously utilized in HRI research[5], these methods are not feasible to use in unrestricted everyday life settings, specifically essential to understand real-world complex human-robot interaction, which contains requirements such as physical contact, non-verbal communication and affective expressions [6]. As a result, the understanding of human social behavior during human-robot interaction remains limited, more so in everyday real-world contexts[2], even though valuable insights from neuroscientific methods could likely be obtained[7].

Social cognition in humans is defined as the perceptual, cognitive, and behavioral processes that are utilized in interpreting others' flexible social behaviors to become adept in social interactions[8]. These processes include Theory of Mind, perception of self and others, socio-emotional processing, and mental attribution[9; 10]. Prior neuroimaging evidence discovered that the medial prefrontal cortex (mPFC) acts as a key area for the cortical network that regulates these social cognition processes in human-human interactions[11; 12; 13]. As an emerging neuroimaging technique, functional near-infrared spectroscopy (fNIRS) has gained more mobility and accessibility over the last decade[14; 15], and can be used for continuous recording in everyday settings, such as participants walking outdoors[16] and pilots in the cockpit flying an aircraft[17]. Therefore, fNIRS as a neurocognitive measure is well-suited to study the social interaction with robots and aligns well with the emerging field of neuroergonomics that aims to study the brain in everyday settings[18].

In this study, we aim to assess the user's social cognition in realistic human-robot interaction settings via neurocognitive measures using functional near-infrared spectroscopy. For this, we designed a study where participants talk and interact face-to-face with Pepper, a humanoid robot (SoftBank Robotics, Figure 1c) by collaborating on a series of naturalistic tasks (Figure 1a). As seen in Figure 1d, the experiment contains three sequential sessions in which the participant interacted robot. In each session, the robot reacted to

the user's responses in one of two ways: either by giving appropriate answers and carrying the interaction as smoothly as possible (congruent attitude); or with the robot making errors, repetitions, or illogical comments to the user's responses (erroneous attitude). Each session lasted approximately 20 minutes and contained three tasks including 4 questions of rapport-building task, followed by 4 questions of island survival and 3 questions of save-the-art tasks[19]. In the rapport-building task, after exchanging introductions, Pepper and the participant made casual conversation by discussing hobbies, music, travel, etc. In the island survival task, the participant chose one of the three items which they thought was most fit for survival on the deserted island. After the selection of an item, Pepper suggested selecting one of the other two items with a reason and waited for a response from the participant for discussion. After the discussion, Pepper asked the participant whether they changed their decision. In the save-the-art task, the participant's task was to rank five different paintings based on their preferences. After ranking, Pepper advised the participant to rank one lower-ranked painting higher and one higher-ranked painting lower, while expressing opinions about why the ranking should be changed, and let the participant respond. Then, Pepper asked the participants whether their ranking was changed or not after the discussion.

Throughout the experiment, the hemodynamic activity over the prefrontal cortex was sampled at 10 Hz by fNIR Devices Model 2000S (Figure 1 a,b). For each participant, light intensity data were filtered with a 100^{th} order finite impulse response low-pass filter with a cut-off frequency of 0.1 Hz to attenuate the high-frequency noise, respiration, and cardiac cycle effects[20]. Then, contamination from motion artifacts is mitigated by the sliding-window-motion-artifact-rejection (SMAR) method[21]. After SMAR, each participant's data were manually inspected for removing any saturation remaining motion artifacts caused by head movement. Oxygenated(HbO) and deoxygenated(HbR) hemoglobin concentration changes for each optode were calculated from the processed light intensity data via the Modified Beer-Lambert Law and were averaged across time for each block. For statistical analysis, average HbO values in each question block were compared using linear mixed models with robot attitude (erroneous vs congruent) as an independent factor.

For this study, we hypothesized that in terms of the hemodynamic response, there will be more brain activity over the mPFC in the congruent attitude condition compared to erroneous condition, as previous studies have shown that users' perceived engagement towards the robot was higher when the robot demonstrated a congruent attitude[22], and mPFC activity was higher during human-human social interactions and collaborations[23; 24]. As shown in Figure 1e, the preliminary fNIRS results from 8 participants revealed that the hemodynamic brain activity over the mPFC area was significantly higher during congruent attitude interaction compared to erroneous attitude.

The preliminary results revealed that when robots are not erroneous, participants showed more activation in brain areas associated with social cognition, which shows consistency with our hypothesis. Albeit the small sample size, this study shows that the assessment of social cognition via neurocognitive measures is a promising approach to advance the HRI field by providing insights from human cognitive processes during unrestricted naturalistic human-robot interactions. These novel insights can benefit the design and development of more congruent and effective social mechanisms in robots in the future.

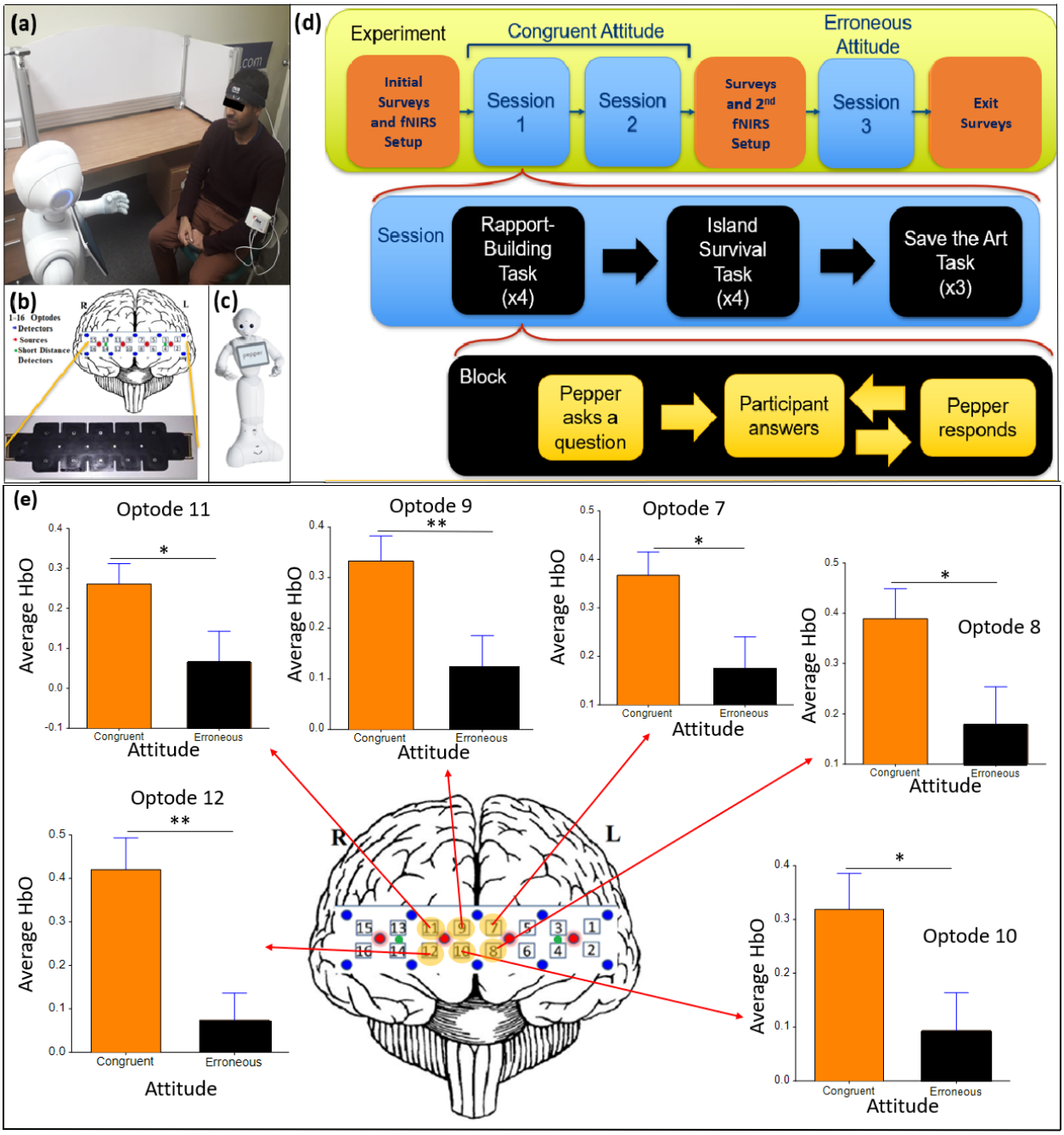

Figure 1. (a) The experiment setup. (b) Detector-source pairs corresponding brain regions for fNIRS headband sensor, which has 16 optodes and 4 LEDs that produces near-infrared light at wavelengths 730 and 850 nm. (c) Pepper, the humanoid robot used in the study. (d) Overall experimental flow and procedure. (e) The preliminary fNIRS results for Robot Attitude condition. In the optodes that represent the mPFC, the brain activity is higher when participants interact with the robot with congruent attitude, specifically optode 7 ($F_{1,250.1} = 5.79$, $p=0.017$), optode 8 ($F_{1,251.2} = 4.68$, $p=0.03$) , optode 9 ($F_{1,251.2} = 6.83$, $p<0.01$) , optode 10 ($F_{1,249.4} = 4.21$, $p=0.04$) , optode 11 ($F_{1,253.0} = 4.84$, $p=0.03$) , optode 12 ($F_{1,255.0} = 9.58$, $p<0.01$).

References:

[1] T.J. Wiltshire, D. Barber, and S.M. Fiore, Towards Modeling Social-Cognitive Mechanisms in Robots to Facilitate Human-Robot Teaming. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 57 (2013) 1278-1282.

[2] A. Henschel, R. Hortensius, and E.S. Cross, Social Cognition in the Age of Human-Robot Interaction. Trends Neurosci 43 (2020) 373-384.

[3] G.-Z. Yang, J. Bellingham, P.E. Dupont, P. Fischer, L. Floridi, R. Full, N. Jacobstein, V. Kumar, M. McNutt, and R. Merrifield, The grand challenges of Science Robotics. Science robotics 3 (2018) eaar7650.

[4] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich, Common metrics for human-robot interaction, Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction, Association for Computing Machinery, Salt Lake City, Utah, USA, 2006, pp. 33–40.

[5] T. Chaminade, D. Rosset, D. Da Fonseca, B. Nazarian, E. Lutcher, G. Cheng, and C. Deruelle, How do we think machines think? An fMRI study of alleged competition with an artificial intelligence. Front Hum Neurosci 6 (2012) 103.

[6] A. Bonarini, Communication in Human-Robot Interaction. Current Robotics Reports 1 (2020) 279-285.

[7] E. Wiese, G. Metta, and A. Wykowska, Robots As Intentional Agents: Using Neuroscientific Methods to Make Robots Appear More Social. Front Psychol 8 (2017) 1663.

[8] R. Adolphs, Social cognition and the human brain. Trends in Cognitive Sciences 3 (1999) 469-479.

[9] C.D. Frith, Social cognition. Philos Trans R Soc Lond B Biol Sci 363 (2008) 2033-9.

[10] J.S. Beer, and K.N. Ochsner, Social cognition: a multi level analysis. Brain Res 1079 (2006) 98-105.

[11] D.M. Amodio, and C.D. Frith, Meeting of minds: the medial frontal cortex and social cognition. Nat Rev Neurosci 7 (2006) 268-77.

[12] J. Hiser, and M. Koenigs, The Multifaceted Role of the Ventromedial Prefrontal Cortex in Emotion, Decision Making, Social Cognition, and Psychopathology. Biol Psychiatry 83 (2018) 638-647.

[13] A.K. Martin, I. Dzafic, S. Ramdave, and M. Meinzer, Causal evidence for task-specific involvement of the dorsomedial prefrontal cortex in human social cognition. Soc Cogn Affect Neurosci 12 (2017) 1209-1218.

[14] R. Parasuraman, and M. Rizzo, Neuroergonomics: The Brain at Work, Oxford University Press, Inc., 2008.

[15] M. Strait, and M. Scheutz, What we can and cannot (yet) do with functional near infrared spectroscopy. Frontiers in Neuroscience 8 (2014).

[16] R. McKendrick, R. Parasuraman, R. Murtza, A. Formwalt, W. Baccus, M. Paczynski, and H. Ayaz, Into the Wild: Neuroergonomic Differentiation of Hand-Held and Augmented Reality Wearable Displays during Outdoor Navigation with Functional Near Infrared Spectroscopy. Front Hum Neurosci 10 (2016) 216.

[17] T. Gateau, H. Ayaz, and F. Dehais, In silico vs. Over the Clouds: On-the-Fly Mental State Estimation of Aircraft Pilots, Using a Functional Near Infrared Spectroscopy Based Passive-BCI. Frontiers in Human Neuroscience 12 (2018).

[18] H. Ayaz, and F. Dehais, Neuroergonomics: The Brain at Work and in Everyday Life, Elsevier Academic Press, 2019.

[19] R. Artstein, D. Traum, J. Boberg, A. Gainer, J. Gratch, E. Johnson, A. Leuski, and M. Nakano, Listen to My Body: Does Making Friends Help Influence People?, FLAIRS Conference, 2017.

[20] H. Ayaz, P.A. Shewokis, A. Curtin, M. Izzetoglu, K. Izzetoglu, and B. Onaral, Using MazeSuite and Functional Near Infrared Spectroscopy to Study Learning in Spatial Navigation. JoVE (2011) e3443.

[21] H. Ayaz, M. Izzetoglu, P.A. Shewokis, and B. Onaral, Sliding-window motion artifact rejection for functional near-infrared spectroscopy, 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, IEEE, 2010, pp. 6567-6570.

[22] M. Salem, F. Eyssel, K. Rohlfing, S. Kopp, and F. Joublin, To Err is Human(-like): Effects of Robot Gesture on Perceived Anthropomorphism and Likability. International Journal of Social Robotics 5 (2013) 313-323.

[23] N. Liu, C. Mok, E.E. Witt, A.H. Pradhan, J.E. Chen, and A.L. Reiss, NIRS-Based Hyperscanning Reveals Inter-brain Neural Synchronization during Cooperative Jenga Game with Face-to-Face Communication. Front Hum Neurosci 10 (2016) 82.

[24] D.D. Wagner, W.M. Kelley, J.V. Haxby, and T.F. Heatherton, The Dorsal Medial Prefrontal Cortex Responds Preferentially to Social Interactions during Natural Viewing. The Journal of Neuroscience 36 (2016) 6917-6925.