Title
Spatio-temporal Neuroimaging of Visual Processing of Human and Robot Actions in Humans

Author
Urgen, Burcu Aysen

Publication Date
2015

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Spatio-temporal Neuroimaging of Visual Processing of Human and Robot Actions in Humans

A dissertation submitted in partial satisfaction of the
requirements for the Doctor of Philosophy

in

Cognitive Science

by

Burcu Ayşen Ürgen

Committee in charge:

Professor Ayse P. Saygin, Chair
Professor Seana Coulson
Professor Anders Dale
Professor Marta Kutas
Professor Piotr Winkielman

2015

The dissertation of Burcu Ayşen Ürgen is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

Chair

University of California, San Diego

2015

DEDICATION

I dedicate this thesis to my family, Belgin, Kamil, Burçin and Buse, who have provided me the best of everything…

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank Professor Ayse P. Saygin for her guidance and support in my academic development as my supervisor during my PhD training.

I also would like to thank Professor Marta Kutas for her mentorship during my PhD years. I would like to thank Professor Howard Poizner for providing EEG set-up in the initial years of my PhD, and Dr. Markus Plank for training me in EEG analysis. I also would like to thank members of Saygin Lab for providing me support and feedback during my PhD years. I especially would like to thank Luke Miller for being a great intellectual friend to discuss fundamental questions in cognitive science.

Chapter 1, in part, is a re-print of the material as it appears in the full conference paper at the Human-Robot Interaction conference, 2015. Urgen, B.A., Li, A.X., Berka, C., Ishiguro, H., Kutas, M. & Saygin, A.P. The dissertation author was the primary investigator and author of this paper.

Chapter 2, in full, is a re-print of the material as it appears in Frontiers in Neurorobotics 2013. Urgen, B.A., Plank, M., Ishiguro, H., Poizner, H., & Saygin, A.P. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in part, is currently being prepared for submission for publication of the material. Urgen, B.A. & Saygin, A.P. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in part, is currently being prepared for submission for publication of the material. Urgen, B.A. & Saygin, A.P. The dissertation author was the primary investigator and author of this paper.

VITA

EDUCATION

PhD in Cognitive Science (2009-2015)
University of California, San Diego (UCSD), CA, USA.
Thesis Title: Spatio-temporal Neuroimaging of Visual Processing of Human and Robot Actions in Humans
Advisor: Ayşe Pınar Saygın, Associate Professor, Dept. of Cognitive Science

MS in Cognitive Science (2007)
Middle East Technical University, Ankara, Turkey.
Thesis Title: A Philosophical Analysis of Computational Modeling in Cognitive Science
Advisor: Samet Bagce, Associate Professor, Dept. of Philosophy
Co-advisor: Cem Bozsahin, Associate Professor, Dept. of Computer Engineering

BS (with Honors) in Computer Engineering and Information Science (2004)
Bilkent University, Ankara, Turkey.

RESEARCH EXPERIENCE AND TRAINING

September 2009-present: Cognitive Neuroscience and Neuropsychology Lab,
University of California, San Diego, La Jolla, CA, USA
Graduate Student Researcher (PI: Ayşe Pınar Saygın)

Summer 2015: Summer Institute in Cognitive Neuroscience, Santa Barbara, CA, USA

Summer 2012: Cold Spring Harbor Laboratory, New York, USA
Computational Neuroscience: Vision - Summer School

August 2010-May 2011: Poizner Lab, Institute for Neural Computation, La Jolla, CA, USA
Graduate Student (PI: Howard Poizner)

March 2010-May 2010: Multimodal Imaging Lab, University of California, San Diego, La Jolla, CA, USA
Graduate Student (PI: Donald J. Hagler)

JOURNAL PUBLICATIONS

Urgen BA, Miller LE (2015). Towards an empirically grounded predictive coding account of action understanding. *Journal of Neuroscience*, 35(12), 4789-4791.

Hofree G, **Urgen BA***, Winkielman P, Saygin AP (2015). Observation and imitation of actions performed by humans, androids, and robots: An EMG study. *Frontiers in Human Neuroscience*, 9:364.

Urgen BA, Plank M, Ishiguro H, Poizner H, and Saygin AP (2013) EEG theta and mu oscillations during perception of human and robot actions. *Frontiers in Neurorobotics.* 7:19.

*** Joint first author.**

MANUSCRIPTS IN PREPARATION

Urgen BA, Saygin AP (in prep). Visual action population codes in the human brain.

Urgen BA, Li A, Ishiguro H, Berka C, Kutas M, Saygin AP (in prep). Predictive processing as a mechanism for uncanny valley: Evidence from electrical brain activity.

Miller LE, **Urgen BA**, Florendo M, Cook J, Saygin AP (in prep). Perception of moving objects: Interactions between (biological) form and motion.

Miller LE, **Urgen BA**, Saygin AP (in prep). Biological motion perception and social cognition: An extended analysis of Miller & Saygin (2013).

Saygin AP, **Urgen BA**, Hagler DJ, and Driver J (in prep). Unattended biological motion drives retinotopy in human lateral temporal cortex.

CONFERENCE PUBLICATIONS

Urgen BA, Li A, Berka C, Kutas M, Ishiguro H, Saygin AP. “Predictive coding and the Uncanny Valley hypothesis: Evidence from electrical brain activity”. Cognition: A Bridge between Robotics and Interaction Workshop at the 10th ACM/IEEE International Conference on Human-Robot Interaction, Portland, Oregon, 2015.

Urgen BA, Plank M, Ishiguro H, Poizner H, Saygin AP. “Temporal Dynamics of Action Perception: The Role of Biological Appearance and Motion Kinematics”. 34th Annual Conference of Cognitive Science Society, Sapporo, Japan, 2012.

Saygin AP, Chaminade T, **Urgen BA**, Ishiguro H. "Cognitive neuroscience and robotics: A mutually beneficial joining of forces". Robotics: Science and Systems (RSS), Human-robot interaction: Perspectives and contributions to robotics from the human sciences, 2011, Los Angeles, CA, USA.

CONFERENCE PRESENTATIONS

Burcu A. Urgen, Ayse P. Saygin**.** Visual action population codes in the human brain. Annual Conference of Organization for Human Brain Mapping, Honolulu, Hawai, 2015.

Burcu A. Urgen, Alvin Li, Chris Berka, Marta Kutas, Ayse P. Saygin. Predictive coding as a mechanism for uncanny valley: Event-related potentials and source localization. Annual Conference of Organization for Human Brain Mapping, Honolulu, Hawai, 2015.

Burcu A. Urgen, Ayse P. Saygin**.** Representational similarity analysis of fMRI responses in brain areas involved in visual action processing. Annual Meeting of Vision Sciences Society, Tampa, Florida, 2015.

Wednesday Bushong, **Burcu A. Urgen**, Luke E. Miller, Ayse P. Saygin. Influence of form and motion on biological motion prediction. Annual Meeting of Vision Sciences Society, Tampa, Florida, 2015.

Burcu A. Urgen, Alvin Li, Hiroshi Ishiguro, Marta Kutas, Ayse P. Saygin. Predictive account of uncanny valley: Evidence from event-related brain potentials. Annual Meeting of Psychonomic Society, 2014.

Burcu A. Urgen, Wayne Khoe, Alvin Li, Ayse P. Saygin. Visual evoked potentials in response to biological and non-biological agents. Annual Meeting of Vision Sciences Society, Tampa, Florida, 2014.

Burcu A. Urgen, Markus Plank, Hiroshi Ishiguro, Howard Poizner, Ayse P. Saygin. EEG Mu and Theta Oscillations during Perception of Human and Robot Actions, Annual Meeting of Society for Neuroscience, San Diego, 2013.

Ayse P. Saygin, **Burcu A. Urgen**, Markus Plank, Thiery Chaminade, James Kilner, Jon Driver, Howard Poizner, Marta Kutas, Chris Frith, Hiroshi Ishiguro. The Visual Processing of Body Movements Studies with Natural and Artificial Agents, Annual Meeting of Society for Neuroscience, New Orleans, 2012.

Burcu A. Urgen, Markus Plank, Marta Kutas, Howard Poizner, Ayse P. Saygin**. "**Temporal Characteristics of Neural Processing During Action Perception: The Role of Biological Form and Biological Motion". 5th Annual Inter-science of Learning Center Student and Post-doc Conference, San Diego, 2012.

Burcu A. Urgen, Markus Plank, Marta Kutas, Howard Poizner, Ayse P. Saygin. "Temporal Characteristics of Neural Processing during Action Perception: The Role of Biological Form and Biological Motion". Annual Meeting of Vision Sciences Society, Naples, Florida, USA, 2012.

Burcu A. Urgen, Markus Plank, Marta Kutas, Howard Poizner, Ayse P. Saygin. “ERP Signatures Associated with Biological Form and Biological Motion Processing During Action Perception”. Annual Meeting of Cognitive Neuroscience Society, Chicago, USA, 2012.

Ayse P. Saygin, Thierry Chaminade, **Burcu A. Urgen**, Hiroshi Ishiguro, Jon Driver, Chris Frith. "The Perception of Body Movements: The Role of Biological Motion and Form". Annual Meeting of Vision Sciences Society, 2011, Naples, Florida, USA.

Burcu A. Urgen, Donald J. Hagler, Jon Driver, Ayse P. Saygin. "Top-down and Bottom-up Modulation of Retinotopic Activity in Temporal and Parietal Cortex". Annual Meeting of Vision Sciences Society, 2011, Naples, Florida, USA.

Burcu A. Urgen, Markus Plank, Howard Poizner, Ayse P. Saygin. "Social Cognition and Interactive Artifical Agents". 10th Anniversary Celebration of California Institute for Telecommunications and Information Technology, 2010, La Jolla, CA, USA.

Ayse P. Saygin, **Burcu A. Urgen**, Donald J. Hagler, Jon Driver. "Retinotopic Maps in Human Cortex: Effects of stimulus structure, attentional load and top-down attention". Annual Meeting of Society for Neuroscience, 2010, San Diego, CA, USA.

Burcu A. Urgen. "A Philosophical Analysis of Computational Modeling in Cognitive Science", Middle East Technical University, Ankara, Turkey (Master Thesis, 2007).

INVITED TALKS

Burcu A. Urgen (2014). Spatio-temporal Neuroimaging of Visual Processing of Body Movements in Humans. Talk at COGS 1 class as a guest lecturer in Spring 2015 at UCSD.

Burcu A. Urgen (2015**).** Spatio-temporal Neuroimaging of Visual Processing of Body Movements in Humans. Talk at Department of Neuroscience, University of Parma, Parma, Italy.

Burcu A. Urgen (2015**).** Spatio-temporal Neuroimaging of Visual Processing of Body Movements in Humans. Talk at the Vision Journal Club, Experimental Psychology, University College London, London, UK.

Burcu A. Urgen (2014)**.** Cognitive Neuroscience and Robotics: An Interdisciplinary Approach for Artificial Agent Design. Talk at COGS 8 - Hands-on Computing class as a guest lecturer, Fall 2014, UCSD.

Burcu A. Urgen (2014). Multivariate Pattern Analysis with fMRI in Cognitive Neuroscience. Talk at Computer Engineering, Middle East Technical University, Ankara, Turkey.

Burcu A. Urgen (2014)**.** Spatio-temporal Neuroimaging of Visual Processing of Body Movements in Humans. Talk at UMRAM, Bilkent University, Ankara, Turkey.

Burcu A. Urgen (2014). Action Recognition in the Human Brain. Talk at COGS1 class as a guest lecturer, Spring 2014 at UCSD.

Burcu A. Urgen (2014). Spatio-temporal Neuroimaging of Visual Processing of Body Movements in Humans. Talk at Pineda Lab, Cognitive Science, Spring 2014 at UCSD.

Burcu A. Urgen (2014). Spatio-temporal Neuroimaging of Visual Processing of Body Movements in Humans. Talk at COGS 11 class as a guest lecturer in Winter 2014 at UCSD.

Burcu A. Urgen (2013). Spatio-temporal Neuroimaging of Visual Processing of Body Movements in Humans. Talk held at the Interdisciplinary Scholar Award Ceremony (as one of the awardees) organized by Graduate Student Association and Office of Graduate Studies, University of California San Diego, USA.

Burcu A. Urgen (2012). EEG Studies of Visual Processing of Body Movements. Talk held at the Swartz Center for Computational Neuroscience, University of California, San Diego, USA.

FELLOWSHIPS, HONORS, AWARDS

2015: Summer Institute in Cognitive Neuroscience Course Fellowship, Santa Barbara, CA, USA
2015: Travel Fund Award, Department of Cognitive Science, UCSD, La Jolla, CA, USA (VSS & HRI, 2015)
2014: Dean of Social Sciences Travel Fund Award, UCSD, La Jolla, CA, USA (VSS Meeting, 2014)
2014: Travel Fund Award, Department of Cognitive Science, UCSD, La Jolla, CA (VSS Meeting, 2014)
2013: Interdisciplinary Scholars Award, UCSD, La Jolla, CA, USA
2012: Cold Spring Harbor Laboratory Summer School Fellowship, New York, USA
2012: Travel Fund Award, Department of Cognitive Science, UCSD, La Jolla, CA, USA (CNS Meeting, 2012)
2011: Graduate Excellence Award, Department of Cognitive Science, UCSD, La Jolla, CA
2011: Friends of the International Center Scholarship, UCSD, La Jolla, CA, USA
2011: Dean of Social Sciences Travel Fund Award, UCSD, La Jolla, CA, USA (VSS Meeting, 2011)
2011: Travel Fund Award, Department of Cognitive Science, UCSD, La Jolla, CA, USA (VSS Meeting, 2011)

2010-2011: Graduate Student Fellowship, California Institute for Telecommunications and Information Technology Fellowship, La Jolla, CA, USA
2009-2013: Graduate Student Departmental Funding, Department of Cognitive Science, UCSD, La Jolla, CA, USA
2004: Graduation in Honor List, Bilkent University, Ankara, Turkey

TEACHING EXPERIENCE

2013 – Instructor, Dept. of Cognitive Science, UCSD, La Jolla, CA, USA.
COGS 119 – MATLAB for Experimental Research

2009 – present: Teaching Assistant, Dept. of Cognitive Science, UCSD, La Jolla, CA, USA.

<u>Undergraduate and graduate courses assisted at UCSD:</u>
COGS 1 Introduction to Cognitive Science
COGS 11 Minds and Brains
COGS 14A Introduction to Research Methods
COGS 102C Cognitive Studio
COGS 107B Systems Neuroscience
COGS 107C Cognitive Neuroscience
COGS 119/219 MATLAB for Experimental Research
COGS 187A Usability and Information Architecture

2005 – 2009: Research and Teaching Assistant, Dept. of Cognitive Science, Informatics Institute, Middle East Technical University (METU), Ankara, Turkey.

<u>Graduate courses assisted at METU:</u>
COGS 501 Formal Languages and Linguistics
COGS 502 Logic and Programming
COGS 511 Computational Models of Mind
COGS 515 Artificial Intelligence for Cognitive Science
COGS 536 Research Methods and Statistics for Cognitive Science
COGS 541 Language Acquisition
COGS 590 Graduate Seminar

WORK EXPERIENCE

June-July 2003: KOÇBANK Computer Center, Çamlıca, İstanbul, Turkey.
Summer Intern (database management systems)

June-July 2002: ASELSAN (Software Company for Military), Macunköy, Ankara, Turkey.
Summer Intern (software engineering)

PROFESSIONAL MEMBERSHIPS

Society for Neuroscience
Vision Sciences Society
Cognitive Neuroscience Society
Women in Cognitive Science
Organization for Human Brain Mapping

PROFESSIONAL SERVICE:

2011-2012: Graduate Student Volunteer, as “Distinguished Speaker Coordinator”, Department of Cognitive Science, University of California, San Diego, La Jolla, CA, USA.

2010-2011: Graduate Student Volunteer, member of “Posters, Brochures, Recruitment Committee”, Department of Cognitive Science, University of California, San Diego, La Jolla, CA, USA.

Ad-hoc reviewer for the journals *Neuroimage*, *Social Cognitive and Affective Neuroscience, Behavioral Brain Research, PLOS One, Frontiers in Psychology.*

LEADERSHIPS AND VOLUNTEER ACTIVITIES

September 2010 - present: Graduate Student Mentor, Cognitive Neuroscience and Neuropsychology Lab, University of California, San Diego, La Jolla, CA, USA
(*Mentored undergraduate and graduate students in EEG data collection and analysis, fMRI data collection and analysis, and behavioral and online experiments on Mechanical Turk*)

Summer 2010, 2011, 2012: Graduate Student Mentor, “Research Experience for High School Students (REHS)” program organized by San Diego Supercomputer Center, University of California, San Diego, La Jolla, CA, USA
(*Mentored high school students from San Diego area in various projects including online experiments on Amazon Mechanical Turk and character animation using 3D animation software Maya and Motion Builder*)

2007 - 2008: Co-founder and co-organizer of student-run journal club “Thinking about Cognition”, Middle East Technical University, Ankara, Turkey.

2001: Instructor, Computer Club of Bilkent University, Ankara, Turkey
(*Lectured and ran lab sessions on "Algorithms and Java Programming"*)

OTHER INFORMATION

Languages

English, Turkish (native), German (basic)

Computer Skills

Programming languages: MATLAB (experienced active user), Java, C/C++, Prolog, Lisp.
Major Software: EEGLAB (EEG analysis tool), ERPLAB (EEG analysis tool), Brain Vision Analyzer (EEG Analysis tool), SPM (Data Analysis tool for Functional Magnetic Resonance Imaging), Freesurfer (automated tools for reconstruction of the brain's cortical surface from structural MRI data), AFNI (Data Analysis tool for Functional Magnetic Resonance Imaging), Princeton MVPA Toolbox, SOAR (Cognitive Modeling tool), ACT-R (Cognitive Modeling tool), Cogent (Cognitive Modeling tool)
Operating systems: Windows, UNIX/Linux, Mac OS

References: Available upon request

ABSTRACT OF THE DISSERTATION

Spatio-temporal Neuroimaging of Visual Processing of Human and Robot Actions in Humans

by

Burcu Ayşen Ürgen

Doctor of Philosophy in Cognitive Science

University of California, San Diego, 2015

Professor Ayse P. Saygin, Chair

Successfully perceiving and recognizing the actions of others is of utmost importance for the survival of many species. For humans, action perception is considered to support important higher order social skills, such as communication, intention understanding and empathy, some of which may be uniquely human. Over the last two decades, neurophysiological and neuroimaging studies in primates have identified a network of brain regions in occipito-temporal, parietal and premotor cortex that are associated with perception of actions, also known as the Action Observation Network (AON). Despite growing body of literature, the functional properties and connectivity patterns of this network remain largely unknown. The goal of this dissertation is to address questions about functional properties and connectivity patterns of AON with a specific focus on whether this system shows specificity for biological agents. To this end, we collaborated with a robotics lab, and manipulated the humanlikeness of agents that perform recognizable actions by varying visual appearance and movement kinematics. We then used a range of measurement modalities including cortical EEG oscillations, event-related brain potentials (ERPs), and fMRI together with a range of analytical techniques including pattern classification, representational similarity analysis (RSA), and dynamical causal modeling (DCM) to study the functional properties, temporal dynamics, and connectivity patterns of the AON. While our findings shed light whether the human brain shows specificity for biological agents, the interdisciplinary work with robotics also allowed us to address questions regarding human factors for artificial agent design in social robotics and human-robot interaction such as uncanny valley, which is concerned with what kind of artificial agents we should design so that humans can accept them as social partners.

INTRODUCTION

Successfully perceiving and recognizing the actions of others is of utmost importance for the survival of many species. For humans, action perception is considered to support important higher order social skills, such as communication, intention understanding and empathy, some of which may be uniquely human (Blake and Shiffrar, 2007). Over the last two decades, neurophysiological and neuroimaging studies in primates have identified a network of brain regions in occipito-temporal, parietal and frontal cortex that are associated with perception of actions, also known as the Action Observation Network (AON) or Mirror Neuron System (MNS) (Rizzolatti and Craighero, 2004; Vangeneugden et al. 2009; 2011; 2014; Caspers et al., 2010; Nelissen et al., 2011; Saygin, 2012a). This system, in its classic formulation, consists of three nodes: the posterior superior temporal sulcus (pSTS), which serves as the visual input to the system by getting visual information from the dorsal and ventral pathways of the visual cortex (Giese and Poggio, 2003); and two sets of regions in the parietal and frontal cortices (Figure A).

Among other regions in the parietal and frontal cortices such as intra-parietal sulcus and dorsal premotor cortex, respectively, AON includes inferior parietal lobule (IPL) and ventral premotor cortex (vPM) that contain neurons that discharge during both action execution and action observation, called mirror neurons, in the non-human primate brain (Rizzolatti and Craighero, 2004).

The initial discovery of mirror neurons in the premotor cortex, which is classically part of the motor system has been influential in theorizing about how primates understand the actions of other individuals. Rizzolatti et al. (2001) proposes that the way primates understand other individuals' actions by means of motor resonance, i.e. by recruiting the same neural resources during action observation as of during action execution and has been taken as supporting evidence for the motor simulation idea in the embodied theories of cognition (Rizzolatti and Craighero, 2004).

One of the most prominent models developed by Kilner et al. (2007a; 2007b) proposes that the AON is a predictive system, following the principles of predictive coding (Friston, 2010). In this framework, in contrast to the classic formulation of the AON, which sees action perception strictly as a feedforward process, Kilner et al. (2007a; 2007b) propose that visual action information is processed throughout the AON by means of the reciprocal connections (i.e., both feedforward and feedback) between the pSTS and parietal cortex, and parietal cortex and the premotor cortex.

Despite a growing body of empirical and theoretical literature, two issues remain unclear:

(1) What are the functional properties of each brain region comprising this network? What kind of information is encoded in each region?

(2) How is the information processed across the brain regions? What is the direction of information flow within the network?

The overall objective of my dissertation research is to address specific questions under the umbrella of the general questions in (1) and (2) by studying the functional properties, temporal dynamics and connectivity patterns in the neural system that underlies visual processing of actions in the human brain using functional magnetic resonance imaging (fMRI), electroencephalography (EEG), and advanced analysis methods from computer science such as multivariate pattern analysis and connectivity. More specifically, the focus of my thesis work is whether this system shows specificity for biological agents, particularly for humans. This question is important because the human brain has evolved with biological agents over generations but we have been introduced artificial agents that can perform recognizable actions like humans that are not biological in nature such as robots in the technology era we live in. How the humans respond to these artificial agents could shed light about the specificity of the human brain to biological agents. This general question is addressed by using state-of-the-art robots, which were provided by a robotics lab (Intelligent Robotics Lab, Osaka University, Japan). We manipulate the human-likeness of observed agents that perform recognizable actions by varying the visual appearance and movement kinematics in a series of neuroimaging experiments using methods with excellent temporal and spatial resolution.

On the other hand, the use of robots as experimental stimuli in our neuroimaging experiments has also allowed us to address some important questions in social robotics and human-robot interaction. As humanoid robots become participants in our lives in areas such as education, healthcare, and entertainment, we need to consider an important issue: How should we design artificial agents so that humans socially accept them and can interact with them successfully? An intuitive approach might be to make the robots as humanlike as possible so that they will be more familiar and tap into neural systems for social cognition that are already well-developed in the human brain. However, increasing humanlikeness does not necessarily result in increasing acceptance from humans (Pollick, 2010). One example is the phenomenon “uncanny valley” first introduced by the robotics researcher Mori (1970). Uncanny valley refers to the idea that humans give positive responses to artificial agents as they have increasing humanlikeness but the relationship between human responses and humanlikeness of the agent will not be a linear one. There is a point where humans start to give negative responses to a set of artificial agents who are almost human-like. Surprisingly, little work has been done in social robotics and human-robot interaction to systematically characterize what kind of artificial agents humans will respond positively and accept as social partners, and what kind of artificial agents could fall into the category of “uncanny”. It has been suggested that the social, cognitive, and neurosciences would be invaluable to identify design principles for the development of artificial agents (MacDorman and Ishiguro, 2006; Saygin et al., 2011; Saygin, 2012b). As such, several studies in this thesis used well-established neural measures to characterize how humans respond to artificial agents that deviate from human norms with the goal of identifying design principles in the development of artificial agents and informing social robotics and human-robot interaction researchers.

In sum, this thesis on the one hand aims to understand how the human brain responds to agents that it did not evolve with over generations and whether those responses are similar to the ones for humans during perception of actions. On the other hand, it aims to demonstrate that decades of cognitive neuroscience research can allow us to address some fundamental questions in social robotics and human-robot interaction and can guide development of design principles for artificial agents who are increasingly becoming participants in the technology era we live in. In short, it attempts to show that artificial agents such as robots can allow us to learn about the human brain but the brain science can also inform social robotics.

The rest of the dissertation is organized as follows. Chapter 1 aims to understand the underlying mechanism of the phenomenon uncanny valley (Urgen et al., 2015).

Theoretical work proposes predictive processing as a possible underlying mechanism for uncanny valley but no empirical work to date has directly tested it. In Chapter 1, we provide evidence that supports this hypothesis using event-related brain potential recordings from the human scalp, which indicate that uncanny valley might be explained by violation of one's expectations about human norms when encountered with very realistic artificial human forms.

Chapter 2 investigates the cortical EEG oscillatory patterns during action observation and the modulation of these patterns by the form and motion of the agent being observed using EEG, time-frequency analysis, and machine learning (Urgen et al., 2013). Our results indicate that mu oscillations over sensorimotor cortex show a similar pattern during observation of human actions and of agents that deviate from human appearance and motion. On the other hand, theta oscillations over frontal cortex show sensitivity to the appearance of the agent observed suggesting that agents that do not look biological enough might results in greater memory processing demands.

Chapter 3 investigates the representational properties of brain regions that are known to be involved in action perception using fMRI and a multivariate pattern analysis technique called representational similarity analysis (Kriegeskorte et al., 2008). Our results suggest that during visual action processing, pSTS pools information from the early visual areas to compute the identity of the agent, and passes that information to regions in parietal and premotor cortex that code higher-level aspects of actions, consistent with computational models of visual action recognition

Chapter 4 investigates the flow of information within the action observation network using fMRI and an effective connectivity technique, dynamical causal modeling

(DCM) (Friston et al., 2003). Specifically, we investigate the influence of two nodes of the AON, pSTS and premotor cortex over the third node, parietal cortex and how this influence is affected by congruence between the appearance and motion of the observed agent. Our DCM results suggest that the influence on parietal node is likely through a feedback connection from premotor cortex during perception of actions that violate people's predictions, supporting the predictive coding account of action perception (Kilner et al., 2007a; 2007b).

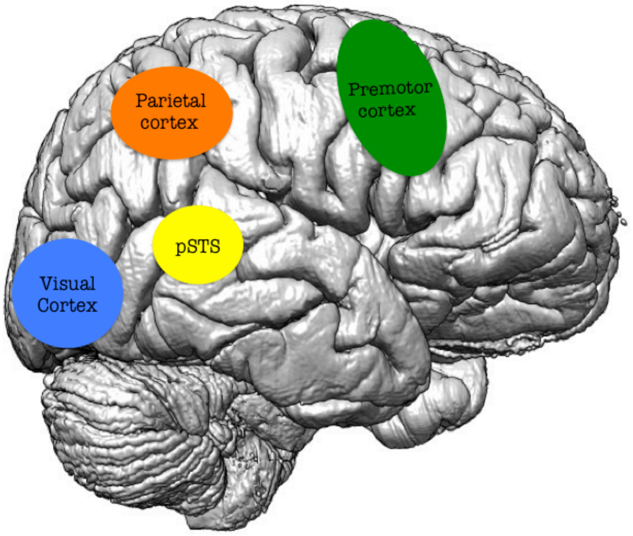

Figure A: Depiction of the Action Observation Network (AON) in the human brain. In its classical formulation, the posterior superior temporal sulcus (pSTS), serves as the visual input to the core nodes of the AON by getting visual information from the dorsal and ventral pathways of the visual cortex; and sends that information to parietal and premotor cortex.

REFERENCES

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology*, *58*, 47–73.

Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *NeuroImage*, *50*(3), 1148–67.

Friston, K.J., Harrison, L. and Penny, W. (2003). Dynamical causal modeling. Neuroimage, *19*(4), 1273-1302.

Friston, K.J. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11, 127–38.

Giese, M.A. & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*, *4*(3), 179–92.

Iacoboni, M. & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience, 7*(12), 942-951.

Kilner, J.M., Friston, K.J., Frith, C.D. (2007a) The mirror-neuron system: a Bayesian perspective. *Neuroreport*, *18*:619-623.

Kilner, J.M., Friston, K.J., Frith, C.D. (2007b) Predictive coding: an account of the mirror neuron system. *Cognitive Processing, 8*:159-166.

Kriegeskorte, N., Mur, M. and Bandettini, P. (2008). Representational Similarity Analysis - Connecting the Branches of Systems Neuroscience. *Frontiers in Systems Neuroscience, 2*(4).

MacDorman, K.F. and Ishiguro, H. 2006. The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies, 7*(3): 297-337.

Mori, M. 1970. The uncanny valley. *Energy*, 7(4): 33-35.

Nelissen, K., Borra, E., Gerbella, M., Rozzi, S., Luppino, G., Vanduffel, W., Rizzolatti, G. & Orban, G. A. (2011). Action observation circuits in the macaque monkey cortex. *Journal of Neuroscience*, *31*(10), 3743–56.

Pollick, F.E. (2010). In search of the uncanny valley. *UCMedia 2009, LNICST 40*, 69-78.

Rizzolatti, G., Fogassi, L. and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience, 2*, 661-670.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–92.

Saygin, A. P., Chaminade, T., Urgen, B.A. and Ishiguro, H. (2011). Cognitive neuroscience and robotics: A mutually beneficial joining of forces. *Robotics: Systems and Science,* Los Angeles, CA.

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. *Social Cognitive Affective Neuroscience, 7*(4): 413-422.

Saygin, A.P. (2012a). Biological motion perception and the brain: Neuropsychological and neuroimaging studies. In K. Johnson & M. Shiffrar (Eds.). *People Watching: Social, Perceptual, and Neurophysiological Studies of Body Perception.* Oxford Series in Visual Cognition, Oxford University Press.

Saygin, A.P. (2012b). What can the Brain Tell us about Interactions with Artificial Agents and Vice Versa? *Workshop on Teleoperated Androids, 34th Annual Conference of the Cognitive Science Society*, Sapporo, Japan.

Urgen, B. A., Plank, M., Ishiguro, H., Poizner, H., & Saygin, A. P. (2013). EEG theta and Mu oscillations during perception of human and robot actions. *Frontiers in Neurorobotics*, *7*, 19.

Urgen, B.A, Li, A.X., Berka, C., Kutas, M., Ishiguro, H., Saygin, A.P. (2015). "Predictive coding and the uncanny valley hypothesis: Evidence from electrical brain activity". *Workshop on Cognition: A Bridge Between Robotics and Interaction* at the *10th ACM/IEEE International Conference on Human-Robot Interaction*, Portland, OR, USA.

Vangeneugden, J., Pollick, F., & Vogels, R. (2009). Functional differentiation of macaque visual temporal cortical neurons using a parametric action space. *Cerebral Cortex*, *19*(3), 593–611.

Vangeneugden, J., De Maziere, P. a., Van Hulle, M. M., Jaeggli, T., Van Gool, L., & Vogels, R. (2011). Distinct Mechanisms for Coding of Visual Actions in Macaque Temporal Cortex. *Journal of Neuroscience*, *31*(2), 385–401.

Vangeneugden, J., Peelen, M. V, Tadin, D., & Battelli, L. (2014). Distinct neural mechanisms for body form and body motion discriminations. *Journal of Neuroscience*, *34*(2), 574–85.

CHAPTER 1

Predictive Processing and Uncanny Valley: Evidence from Electrical Brain Activity

Abstract

Uncanny valley refers to humans' negative reaction to almost-but-not-quite-human agents. Theoretical work proposes predictive processing as a possible underlying mechanism for uncanny valley but no empirical work to date has directly tested it. Here, we provide evidence that supports this hypothesis using event-related brain potential recordings from the human scalp, which indicate that uncanny valley might be explained by violation of one's expectations about human norms when encountered with very realistic artificial human forms.

1.1 INTRODUCTION

Our social milieu has changed tremendously in recent years, introducing us to social partners that are dramatically different from those the human brain has evolved with over many generations. Specifically, from guiding students in learning math and science, to helping children with autism and stroke survivors in their exercises, artificial human forms such as robots are rapidly becoming participants in our lives. As humanoid robots become participants in our lives, one important issue to consider is: How should we design artificial agents so that humans socially accept them and can interact with them successfully? An intuitive approach might be to make the robots as humanlike as possible so that they will be more familiar and tap into neural systems for social cognition that are already well-developed in the human brain.

However, increasing humanlikeness does not necessarily result in increasing acceptance (Pollick, 2010). Uncanny valley is a phenomenon that refers to humans' response to artificial human forms, which possess almost human-like characteristics. In describing the phenomenon, Mori (1970), who introduced the term, proposes that the relationship between humanlikeness and humans' response to non-human agents is not a linear one. According to his framework, the increasing humanlikeness of an agent elicits positive responses from humans up to a certain point, where increasing humanlikeness starts to elicit negative responses, which forms a deep valley (Figure 1.1). Furthermore, it has been suggested that if the agent is moving, the responses will be much more pronounced compared to static form of the agent.

As humanoid robots became more feasible to develop in recent years, the uncanny valley became a frequently discussed issue from both a theoretical and a practical viewpoint. MacDorman and Ishiguro (2006) contended that the social, cognitive, and neurosciences would be invaluable if we were to understand this intriguing phenomenon. Indeed, empirical studies have recently been exploring the anecdotally well-known, but scientifically uncharted valley. Several subjective rating studies with a range of humanoid stimuli claimed that the uncanny valley might be a legitimate psychological phenomenon: For example, MacDorman et al. (2009), Flach et al. (2012), and Seyama and Nagayama (2007) used computer-animated faces and asked human subjects to rate such dimensions as humanlikeness, eeriness, attractiveness or pleasantness. In a similar fashion, Poliakoff et al. (2013) recently used human, robot, and prosthetic hand stimuli and reported eeriness ratings that were broadly compatible with the hypothesis. Piwek et al. (2014) also reported evidence for the hypothesized valley for very realistic humanoid agents in a study that collected social acceptability ratings with full-body computer animated agents as stimuli. However, their data did not support Mori's proposal that there would be a more pronounced effect with moving stimuli. More broadly, some studies did not reveal evidence for Mori's hypothetical curve with experimental data. Thompson et al. (2011) varied several motion parameters and explored how they influenced humanlikeness, familiarity, and eeriness ratings of human avatars, and did not find results resembling the hypothesized uncanny valley. The inconsistencies between the studies may be due to different dependent measures that were used in the ratings (e.g., likeability is a complex measure that is correlated with the humanlikeness dimension that is used as the x-axis of the Mori graph (Bartneck et al., 2009; Ho and MacDorman, 2008).

In addition to rating studies, researchers have attempted to use less explicit measures such as gaze behavior to characterize the uncanny valley. Using eye-tracking and a parametrically varying set of avatar faces, Cheetham et al. (2013) showed that ambiguous avatar faces (i.e., those that are at the category boundary between human and avatar) required greater depth of processing in the eye and mouth regions compared with unambiguous avatar faces. Similarly, Matsuda et al. (2012) found that by 9-10 months of age, infants looked longer to highly familiar and strange faces compared with morphed faces near the boundary these categories. Furthermore, Steckenfinger and Ghazanfar (2009) reported that that monkeys looked longer at faces of monkey-like agents that were either of their own species or unrealistic animations compared with very realistic animations. Although this suggests the uncanny valley has earlier evolutionary origins, Lewkowicz and Ghazanfar (2012) used analyses of infants' gaze behavior that early exposure to typical human faces in development constrained uncanny valley-like responses.

Although these and similar recent studies have been a good step to scientifically characterize the uncanny valley, the underlying mechanism remains unclear. Possible mechanisms that have been proposed include threat or disease avoidance, mate selection, and Bayesian estimation or predictive coding hypotheses (MacDorman et al., 2009; Moore, 2012; Saygin et al., 2012). The latter hypothesis is linked to a more general description of neural computational properties of the brain (Rao and Ballard, 1999; Friston, 2010), and therefore promises a scientifically testable framework. According to predictive coding, the uncanny valley is related to expectation violations in neural computing when the brain encounters almost-but-not-quite-human agents. A growing body of work has linked Mori's hypothetical curve to the processing of conflicting perceptual or cognitive cues, varying whether the stimuli are compatible with the elicited expectations or are in violation of them (Ho and MacDorman, 2008; Cheetham et al., 2013; Saygin et al., 2012; Mitchell et al., 2011; Cheetham et al., 2011; Nie et al., 2012; Tinwell et al., 2013; Yamamoto et al., 2009).

Behavioral studies alone are insufficient to directly test and identify mechanisms that underlie the uncanny valley, or to distinguish between alternative theories for numerous reasons. First, dependent measures such as subjective ratings require overt responses, whereas the uncanny valley phenomenon might be better studied with covert responses of humans' subjective states that may even occur outside of awareness (cf. Li et al., 2015). In addition, it is difficult to characterize a complex phenomenon with a single measurement such as pleasantness, familiarity, or eeriness as each can imply different cognitive and emotional states, and captures uncanny valley curve for different robot characteristics (Rosenthal-von der Pütten and Krämer, 2014). Second, behavioral studies only provide the output of the system, and do not address what kind of information processing underlies the phenomenon. Although methods such as eye-tracking or automatic attention paradigms have advantages over rating studies in this respect, to provide a mechanistic account of the uncanny valley, methods from social and cognitive neuroscience are likely to be more fruitful (Steckenfinger and Ghazanfar, 2009; Saygin et al., 2011; Saygin, 2012; Cheetham and Jancke, 2013; Urgen et al., 2013).

Neuroscience methods such as neuroimaging have advantages that can help “demystify” the uncanny valley. First of all, there is the potential to provide valid dependent measures that can be used to operationalize the uncanny valley, and to situate it as part of a cognitive domain. Decades of cognitive neuroscience research have informed us about the basic functions of the human brain, perception and social cognition. It would be fruitful to use accumulated knowledge in these areas to inform robotics about how humans respond to and interact with social agents (MacDorman and Ishiguro, 2006; Saygin et al., 2011; Saygin, 2012). Second, neuroimaging does not require overt responses since brain activity can be monitored on an ongoing basis. Last but not least, neuroimaging provides a rich a set of data, which can be more informative than individual behavioral measures. Temporally sensitive methods in particular provide a means to understand the time course of processing in comparison to ratings or reaction times, which only provide the output of the system. Overall, neuroimaging research has the potential to reveal the underlying mechanisms of the uncanny valley phenomenon.

Indeed, there is now growing interest in using neuroimaging in the field. Cheetham et al. (2011) used functional magnetic resonance imaging (fMRI) with a face stimulus set along a human-avatar continuum and found that ambiguous faces at the category boundary of human and avatar are processed differentially than the unambiguous faces within each category. Behavioral studies with animated faces support this category conflict explanation for uncanny valley, which is in line with the prediction error hypothesis (Moore, 2012; Burleigh et al., 2013; Yamada et al., 2013). Another example is a previous work from our lab (Saygin et al., 2012), which used fMRI and based on the results, proposed predictive coding as a framework for future studies on the uncanny valley and the underlying mechanisms. In this study, brain responses to body movements of agents of varying degrees of humanlikeness with and without conflicting perceptual cues was compared. The agents were a human with biological appearance and biological motion, a very human-like robot (referred to as android) with biological appearance but mechanical motion, and a less human-like robot with mechanical appearance and mechanical motion (Figure 1.2A). Notably, neural activity, especially in the parietal cortex differentiated the android from the other two agents. The data suggested, based on the functional properties of this brain region in the social cognition network, that the uncanny valley might be related to the violation of the brain's internal predictions due to conflicting perceptual cues (appearance and motion). The human and robot agents exhibited congruent appearance and motion profiles (i.e. human looks biological, moves biologically; robot looks mechanical, moves mechanically) whereas the android exhibited incongruent appearance and motion (looks biological but moves mechanically). For the latter agent, the human appearance would elicit predictions that the motion will also be; when the agent instead moves mechanically, the brain network processing the agent would show evidence of processing the violations. The differential activity measured in parietal cortex could reflect this prediction error (Kilner et al., 2007a, 2007b; Saygin et al., 2012).

Although Saygin et al. (2012) used neuroimaging to situate the uncanny valley phenomenon in the scientific context of violation of predictions and the predictive coding theory of neural computations, the study was not a priori designed to test this theory. Thus, the proposed framework ideally needs to be further validated with independent experiments. Furthermore, fMRI has methodological limitations, most notably due to its limited temporal resolution.

Electroencephalography (EEG) is an alternative neuroimaging method that allows recording brain activity with electrodes located on the scalp with excellent temporal resolution. Importantly, a specific dependent measure derived from EEG, the N400 event-related potential (ERP) component, is an ideal measure with which to test the prediction error hypothesis or the uncanny valley. N400 is a negative-going ERP, which peaks around 400 ms after stimulus onset, and is maximal in fronto-central regions of the human scalp (Kutas and Federmeier, 2011). Although the N400 is elicited in response to any meaningful stimulus, its amplitude is greater for semantically or contextually anomalous stimuli (i.e., items that violate expectations).

In the present study, we used well-controlled stimuli, which did and did not violate form-motion predictions, together with electroencephalography (EEG) and a remarkable biomarker of human information processing, the event-related brain potential

N400 component to directly test this theory. We hypothesize that upon exposure to a human-like form, our brains predict human-like behavior, specifically human-like (biological) movement based on a life-long experience of conspecifics. Uncanny valley occurs when those predictions are not met, specifically when encountered with agents having human-like forms but non-human-like movements. No empirical work to date has tested this theory of predictive processing directly.

We presented agents of varying humanlikeness in static and dynamic forms as we recorded EEG. The stimuli consisted of a real human agent with human-like appearance and motion (Human), a realistic robot agent with human-like appearance and non-human-like motion (Android), and a mechanical robot with non-human-like appearance and motion (Robot) (Figure 1.2A). In this stimuli set, the real and mechanical agents (Human and Robot) had *congruent* appearance and motion whereas the realistic agent (Android) had *incongruent* appearance and motion. We hypothesized that the realistic agent (Android) would elicit a greater N400 response in dynamic form than the static form as its human-like form would lead to the prediction that it would move in a human-like way but when it did not, it would violate the brain's predictions. On the other hand, we hypothesized that the N400 amplitude for the static and dynamic forms would not differ for Human and Robot since both possess appearance-motion congruence (Human looks human-like, moves in a human-like way; Robot looks non-human-like, moves in a non-human way). Such a pattern of activity would provide direct empirical evidence for the predictive account of uncanny valley.

1.2 MATERIALS AND METHODS

1.2.1 Participants

Twenty right-handed adults (10 females; mean age = 23.8; SD = 4.8) from the student community at University of California, San Diego participated in the study. They had normal or corrected-to-normal vision, and no history of neurological disorders. Informed consent was obtained in accordance with the university's Human Research Protections Program. Participants were paid $8 per hour or received course credit. One subject's data was excluded due to high noise during EEG recording.

1.2.2 Stimuli

Stimuli consisted of video clips of actions performed by the humanoid robot Repliee Q2 (in Robotic and Human-like appearance) and by the human 'master', after whom Repliee Q2 was modeled (Figure 1.2A, also see Saygin et al. 2012 and Urgen et al. 2013 for details about the stimuli).

We refer to these agents as the Robot, the Android (realistic robot), and the Human conditions. Note that the former two are in fact the same robot. Repliee Q2 has 42 degrees of freedom and can make face, head and upper body movements (Ishiguro, 2006). However, the robot's movements did not match the dynamics of biological motion; it is mechanical or "robotic". The same body movements were videotaped in two appearance conditions. For the Robot condition, Repliee Q2's surface elements were removed to reveal its wiring, metal arms and joints, etc. The silicone 'skin' on the hands and face and some of the fine hair around the face could not be removed but was covered. The movement kinematics for the Android and Robot conditions was identical, since these conditions comprised the same robot, carrying out the very same movements. For the Human condition, the female adult whose face was molded and used in constructing Repliee Q2 was videotaped performing the same actions. She was asked to watch each of Repliee Q2's actions and perform the same action naturally. All agents were videotaped in the same room with the same background. Video recordings were digitized, converted to grayscale and cropped to 400x400 pixels. Videos were clipped such that the motion of the agent began at the first frame of each video.

1.2.3 Procedure

Since prior knowledge can affect judgments of artificial agents differentially (Saygin and Cicekli, 2002), each participant was given exactly the same introduction to the study and the same exposure to the videos. Before starting EEG recordings, participants were shown each video and told whether each agent was a human or a robot, and the name of the action. Participants went through a practice session before the experiment. EEG was recorded as participants watched the images or video clips of the three agents performing eight different upper body actions (drinking from a cup, examining an object with hand, handwaving, turning the body, wiping a table, nudging, introducing self, and throwing a piece of paper). The videos were presented in two modes that we call *motion alone* and *still-then-motion*. In the *motion-alone* condition, 2-second videos were presented. In the *still-then-motion* condition, the first frame of the video was presented for 600-1000 ms (with a uniform probability jitter), and then the full video was played. The experiment consisted of 15 blocks. In each block, the eight videos of each agent were presented once in the *motion-alone* condition, and once in the *still-then-*

motion condition. Stimuli were presented in a pseudo-randomized order ensuring that a video was not repeated on two consecutive trials. Each participant experienced a different pseudo-randomized stimuli sequence.

Stimuli were displayed on a 19” Dell Trinitron CRT monitor at 90 Hz using Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). To prevent an augmented visual evoked potential at the beginning of video onset that might occlude subtle effects between conditions, we displayed a gray screen with a white fixation cross before the start of the video clip or still frame on each trial. Participants were instructed to fixate the fixation cross at the center of the screen for 900-1200 ms (with a uniform probability jitter). A comprehension question was displayed every 6-10 trials, asking participants a true/false question about the action in the just seen video (e.g. Drinking?), after which they responded with a manual key press (Yes/No response).

1.2.4 EEG Recording and Data Analysis

EEG was recorded at 512 Hz from 64 ActiveTwo Ag/AgCl electrodes (Brain Vision, Inc.) following the International 10/20 system. The electrode-offset level was kept below 25 k-Ohm. Two additional electrodes were placed above and below the right eye to monitor oculomotor activity (1 additional electrode was placed on the forehead as a ground of the eye electrodes). The data were preprocessed with MATLAB and the EEGLAB toolbox (Delorme and Makeig, 2004). Each participant’s data were first high-pass filtered at 1 Hz, low-pass filtered at 50 Hz, and re-referenced to average mastoid electrodes behind the right and left ear. Then the data were epoched ranging from 200 ms preceding video or first frame onset to 700 ms after video onset, and were time-locked to the onset of the video clips (*motion-alone* condition, see Procedures) or the first frame (*still-then-motion* condition, see Procedures) to compare the motion and still forms of the agents (we refer to these as *motion* and *still* conditions). Atypical epochs of electromyographic activity were removed from further analysis by semi-automated epoch rejection procedures (kurtosis and probability-based procedures with standard deviation = 6). After preprocessing, grand average event-related brain potentials (ERP) and scalp topographies were computed and plotted for each condition using Brain Vision Analyzer.

1.2.5 Statistical Analysis

The time window between 370-600 ms was considered for N400 analysis based on the grand average ERPs across all conditions. The area under curve measure was used to extract the N400 values for each agent under both motion and still condition for each subject in frontal channels (AF3, AFz, AF4, Fz, F1, F2, F3, F4, F5, F6) since N400 for pictorial stimuli has a more frontal distribution (Kutas and Federmeier, 2011). After preprocessing, data were exported to ERPLAB (http://erpinfo.org/erplab) and area under curve measures were extracted by using this toolbox. We then applied paired t-tests on the average frontal channel activity to compare the motion and still conditions for each agent (Robot, Android, Human). Since we expected motion condition to be greater than the still condition for Android (and no effect for Human and Robot), our t-tests were one-tailed.

1.2.6 Localization of EEG Activity

For identifying the neural generators (sources) of the activity during the N400 period, we used the LORETA method (Pascual-Marqui et al., 1994). LORETA estimates the distributed neural activity in the cortex based on the scalp measurements of ERP differences. Localization of the EEG activity was as follows: First, we computed the N400 differences between the static and motion conditions of each agent (Robot, Android, Human), and then we took the grand average of the N400 differences. We then applied LORETA to the N400 difference waveform in the time interval between 370-600 ms to estimate the distributed neural activity underlying N400.

1.3 RESULTS

1.3.1 N400 component

Our ERP results indicate that observation of all agents elicited an N400 component regardless of the presentation mode (static or dynamic) in frontal sites (AF3, AFz, AF4, Fz, F1, F2, F3, F4, F5, F6) on the human scalp (Figure 1.2B shows ERPs on a representative frontal channel Fz). The amplitude of N400 (measured with area under curve between 370-600 ms averaged across all frontal sites) in the dynamic form was significantly greater than the static form for Android ($t(18) = 2.401$, $p<0.05$), whereas static and dynamic forms did not differ neither for Robot ($t(18) = 0.388$) nor for Human ($t(18) = -0.346$) (Figure 1.2B for ERPs, Figure 1.2C for bar graphs for the amplitude of N400, and Figure 1.3 for scalp maps to see the distribution of the effect on the whole scalp).

1.3.2 Source Analysis

Our source analysis with LORETA suggests that the generator of the N400 component is a widely distributed network including the middle and superior temporal areas, temporal-parietal junction, and prefrontal areas (Figure 1.4). These areas align with the network that has been implicated for N400 with intracranial recordings and MEG in humans (Kutas and Federmeier, 2011). More interestingly, the maximal source density of this network was identified as a region within the inferior parietal lobule (Brodman area 40; x = -59, y = -32, z = 29, MNI coordinates) (Figure 1.4), which is the same area that differentiated the agent with appearance-motion incongruence (Android) from the other agents with appearance-motion congruence (Human and Robot) in an independent fMRI study from our lab with excellent spatial resolution (Saygin et al., 2012). Separate source analysis for each condition during the N400 interval (370-600 ms) also resulted in similar networks (Figure 1.5).

1.4 DISCUSSION

In the present study, we tested the predictive coding as a potential underlying mechanism of the uncanny valley phenomenon. To this end, we used an established method from cognitive neuroscience, namely EEG, specifically focusing on the event related brain potential N400, which has been reliably associated with violation of predictions. Our stimuli consisted of three agents that had different levels of humanlikeness in appearance and motion dimensions: a human agent with biological appearance and motion, an android with biological appearance and mechanical motion, and a robot with mechanical appearance and motion. In this design, human and robot agents exhibited congruence in appearance and motion dimensions, whereas android agent had incongruity in appearance and motion. The agents were presented both still and moving, with the hypothesis that the android condition would elicit an N400 differential (motion-still difference) due to its incongruent appearance and motion, whereas human and robot agents would not, as they had congruent appearance and motion. Our results confirmed these predictions: the moving android elicited a greater N400 component than the still android, whereas no difference was found for the moving and still presentations of the human and robot agents. Thus, these results provides support for the hypothesis that uncanny valley might involve the violation of the brain's predictions.

Our study demonstrates the benefit of using neural dependent measures in testing hypotheses about uncanny valley, whose underlying mechanism has remained unknown. Previous research mainly has focused on behavioral ratings in studying the uncanny valley. While these efforts have been a good step to operationalize the uncanny valley, they fall short for a number of reasons. For one thing, these studies generally ask for an explicit (or conscious) response in a certain dimension such as humanlikeness, eeriness, or familiarity. However, explicit measures might be too restrictive and might not be sufficient to characterize the reaction of the human subjects for uncanny stimuli. Neuroimaging has the advantage to measure human responses implicitly without asking for a specific response. In the present study, N400 was used as such an implicit measure. For another thing, behavioral measures provide only the output of the system (one data point), which is not very informative about the processing stages. Neuroimaging provides a rich set of data, and especially the temporally sensitive methods such as EEG allows one to monitor the information processing during stimulus presentation. In addition, well-established dependent measures, such as N400 as used in this study, help one to situate the uncanny valley in a well-studied cognitive domain.

The use of event-related brain potentials in the present study is complementary to our previous neuroimaging study of action perception that used fMRI with the same stimuli. Saygin et al. (2012) found differential activity in parietal cortex for the android compared with the human and robot, which was interpreted as supporting evidence for the hypothesis that uncanny valley might be due to the incongruity of appearance and motion in the action processing network. The N400 effect for the android in the present study corroborates this interpretation. Using EEG has allowed us to link the uncanny valley phenomenon to cognitive processing using the well-established dependent measure N400.

The current study has broader implications for future research on characterizing the uncanny valley. First of all, the appearance-motion incongruence presented in the current study is one specific violation of one’s predictions. In fact, Mitchell et al. (2011) has shown that conflicting visual and auditory cues (appearance and voice) increase eeriness ratings in evaluating human and robot agents. In another study, Nie et al. (2012) showed that incongruent appearance and touch (non-humanlike appearance and humanlike touch) resulted in fear from the robot in human subjects. Based on their exploratory rating study with a number of robot videos, Ho and MacDorman (2008) suggest that uncanny valley effect might be seen for agents that have mismatches in a variety of dimensions. On the other hand, robots differ from humans in a variety of ways, not only in their physical properties such as appearance, motion, and voice, but also in the way they accomplish tasks. In fact, Yamamoto et al. (2009) showed that children of 2-3-year-old showed different behavioral patterns based on the congruity of the robot appearance and contingency of its behavior upon their own behavior. Similarly, it has been suggested that congruity of facial expressions and actions determines uncanny valley responses (Tinwell et al., 2013). Thus, creating broader range of violations of one's predictions could allow us to understand the sensitivity of humans to deviation from human dimensions.

In addition, individual differences are a recent highlighted aspect of uncanny valley (Walters et al., 2008; Macdorman and Entezari, in press) suggesting that people may show different patterns of reactions to humanoid robots, and this could well be studied with neural dependent measures. The role of experience and learning can also be studied, by testing people who have been exposed to robots or animated characters compared to those who have not (Chen et al., 2011). Previous research suggests that culture (e.g., western vs. eastern) as well as context are important factors in perceiving and interacting with humanoid robots (Oyedele et al., 2007). Thus, these different groups' expectations from robots might differ considerably from each other, resulting in different brain activity patterns in general.

In conclusion, our study demonstrates that studying the uncanny valley with neuroscience methods can help us not only understand the underlying mechanisms but it can also inform human robot interaction. Furthermore, the uncanny valley serves as a window into human social cognition. Studying human brain responses when viewing robots has allowed us to study the functional properties of the neural systems that underlie agent and action processing, which are the building blocks for social interaction (MacDorman and Ishiguro, 2006; Saygin et al., 2011, 2012; Saygin, 2012; Urgen et al., 2013). On the other hand, our study also gives insights about the design parameters of robots that will interact with humans. Based on the current data and emerging underlying mechanisms of the uncanny valley, we suggest the human brain's expectations from a human-like agent should be considered in the design process for successful human-robot interaction. More broadly, we demonstrate that interdisciplinary work can not only improve our understanding of human-robot interaction, but also make individual contributions to both neuroscience and robotics. Just as perceptual illusions using artificial stimuli inform us about underlying mechanisms of perception (Gregory, 1980), studying uncanny valley using artificial agents serves as a window into human social cognition.

ACKNOWLEDGEMENTS

Chapter 1, in part, is a re-print of the material as it appears in the full conference paper at the Human-Robot Interaction conference, 2015. Urgen, B.A., Li, A.X., Berka, C., Ishiguro, H., Kutas, M. & Saygin, A.P. The dissertation author was the primary investigator and author of this paper. This research was supported by NSF (CAREER BCS1151805), DARPA, Kavli Institute for Brain and Mind and the Qualcomm Institute (Calit2). We thank the co-authors Alvin Li, Chris Berka, Marta Kutas, Hiroshi Ishiguro, and Ayse Saygin. We also thank Intelligent Robotics Laboratory at Osaka University for the preparation of the stimuli, Akila Kadambi, Wayne Khoe, Edward Nguyen and Markus Plank for assistance in data collection and analysis.

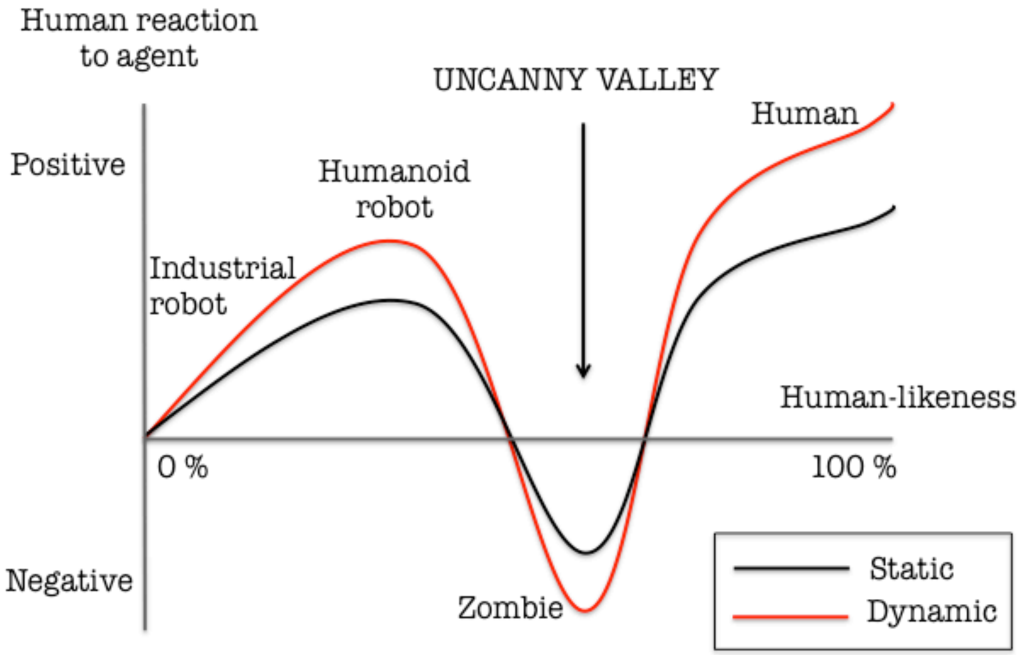

Figure 1.1: Hypothetical curves that depict the uncanny valley effect for static and moving agents in varying levels of humanlikeness.

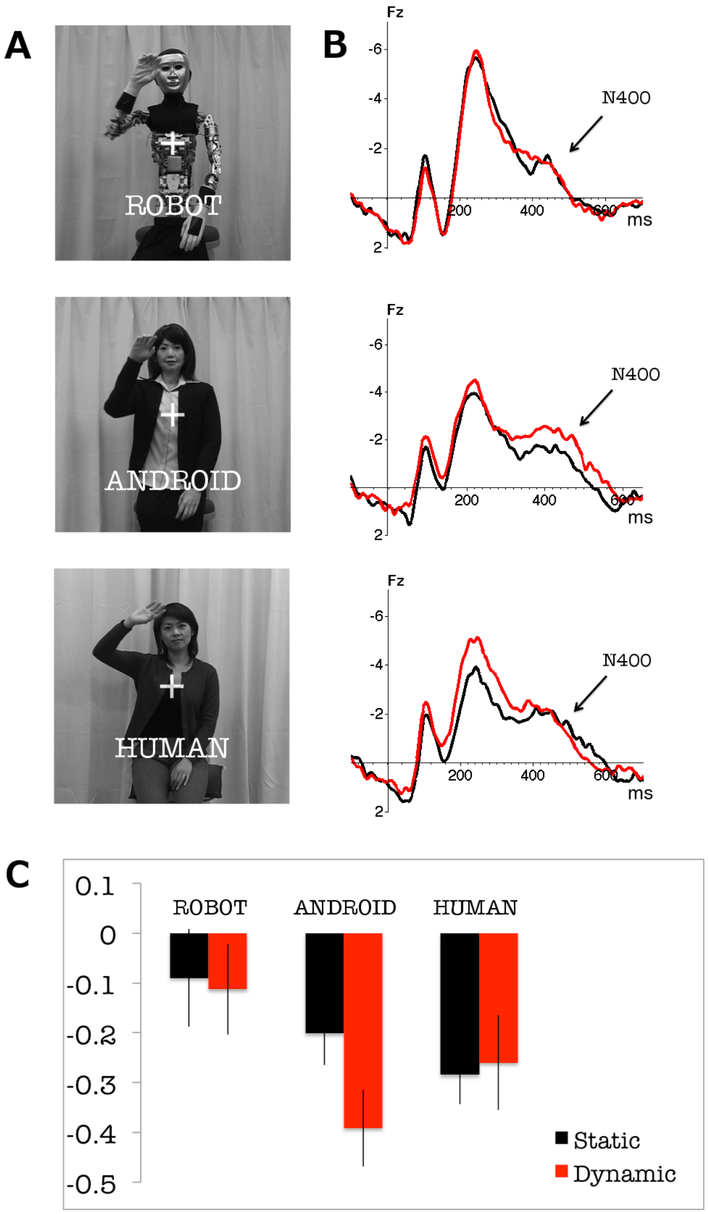

Figure 1.2: Stimuli used in the ERP experiment, ERP plots for the N400 effect, and bar plot for the N400 effect. (A) Sample static frames from the movies used in the EEG experiment depicting the three agents: Robot, Android, Human. (B) ERP plots of a representative frontal site (Fz) for static (black line) and dynamic (red line) forms for each agent (Robot, Android, Human). N400 is greater for moving Android compared to static, whereas no such difference was found for Human or Robot. (C) Bar graphs representing the area under curve for N400 (370-600 ms) for each of the conditions. N400 is significantly greater for dynamic than static form for Android, whereas they did not differ for Robot or Human.

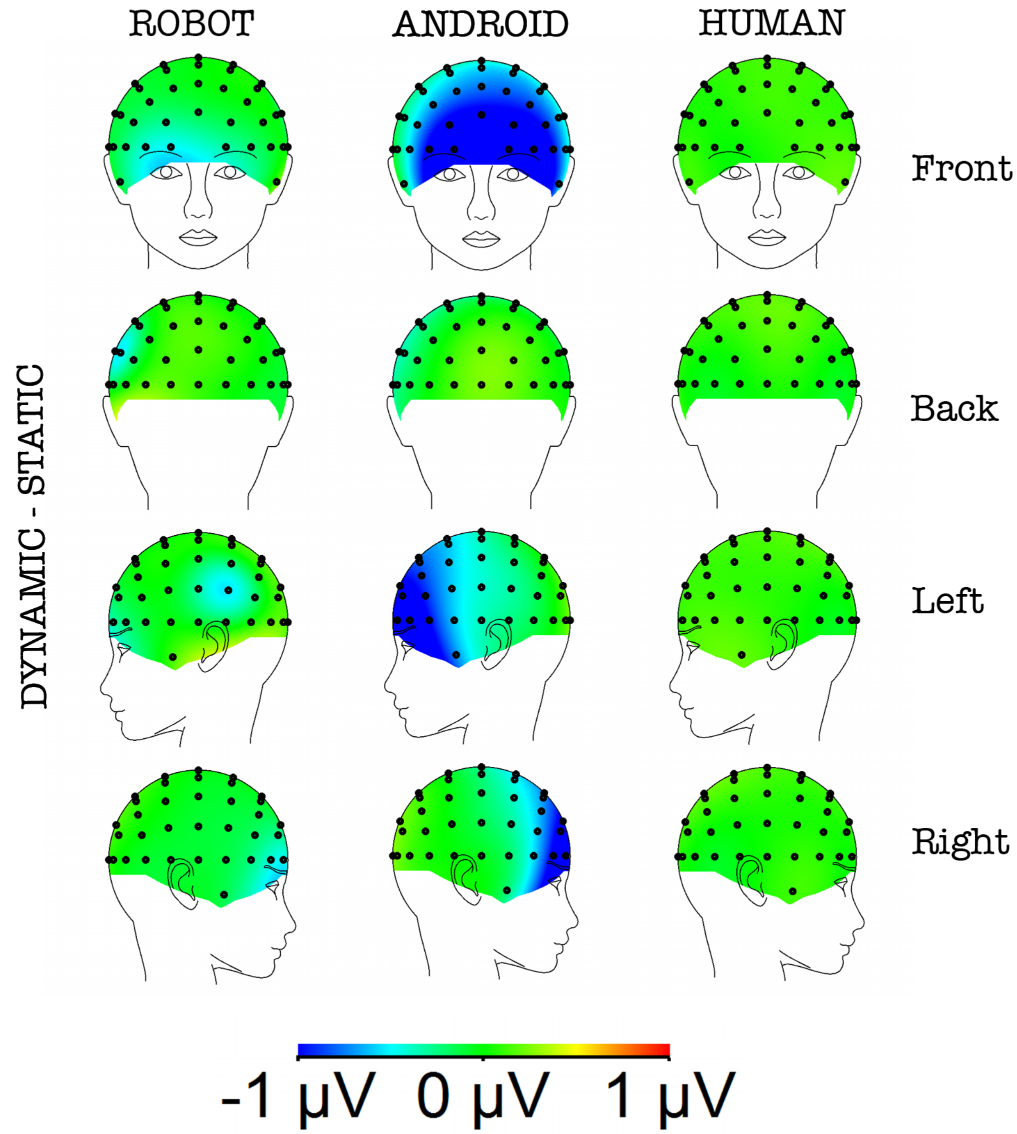

Figure 1.3: ERP scalp maps representing the difference between static and dynamic forms for each agent (Human, Android, Human) in the time interval of the N400 (370 ms – 600 ms).

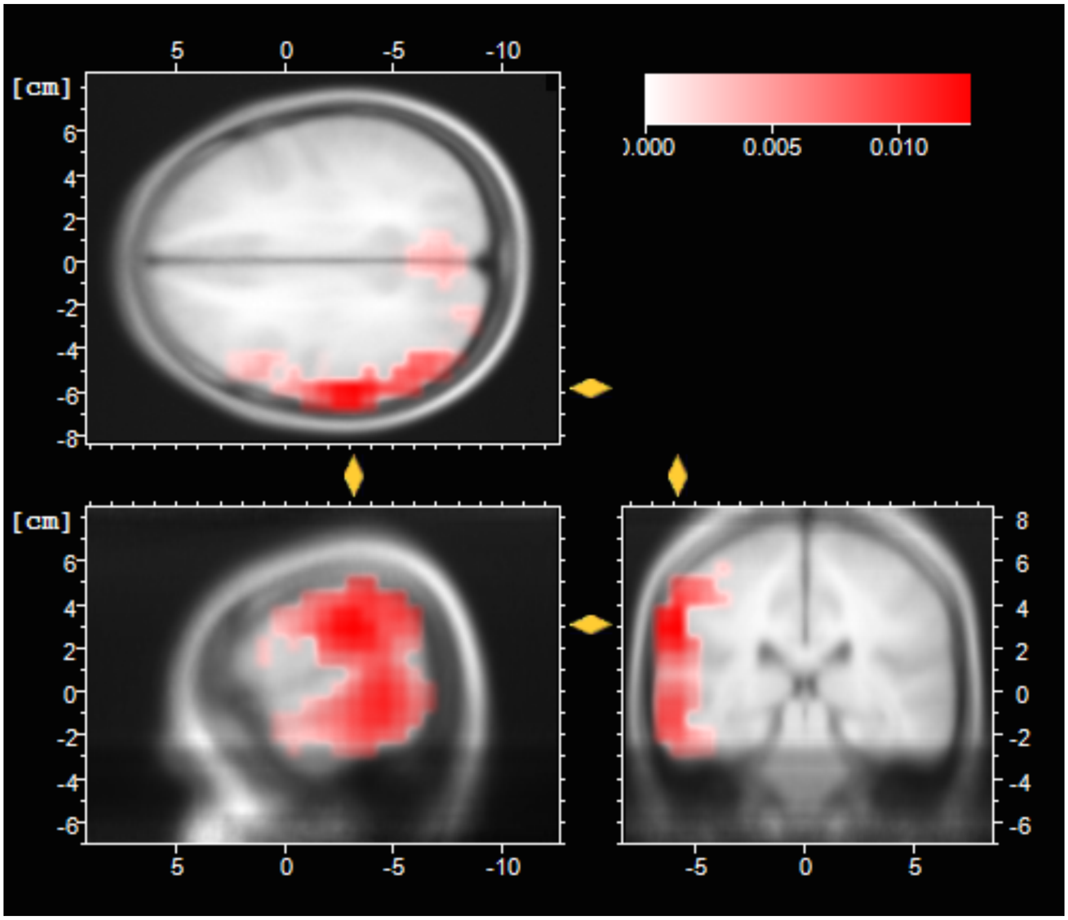

Figure 1.4: LORETA analysis in the N400 (370-600 ms) interval, all conditions collapsed. LORETA analysis for N400 identified a distributed brain activity including middle and superior temporal areas (MTG and STG), tempora-parietal junction (IPL), and frontal areas (IFG, MFG, Medial FG) (Difference waves of all conditions (Robot, Android, Human) are collapsed). The maximal source density of this network is Inferior parietal lobule (IPL) whose coordinates (x = -59, y = -32, z = 29, MNI coordinates) are depicted by the yellow diamond in the figure.

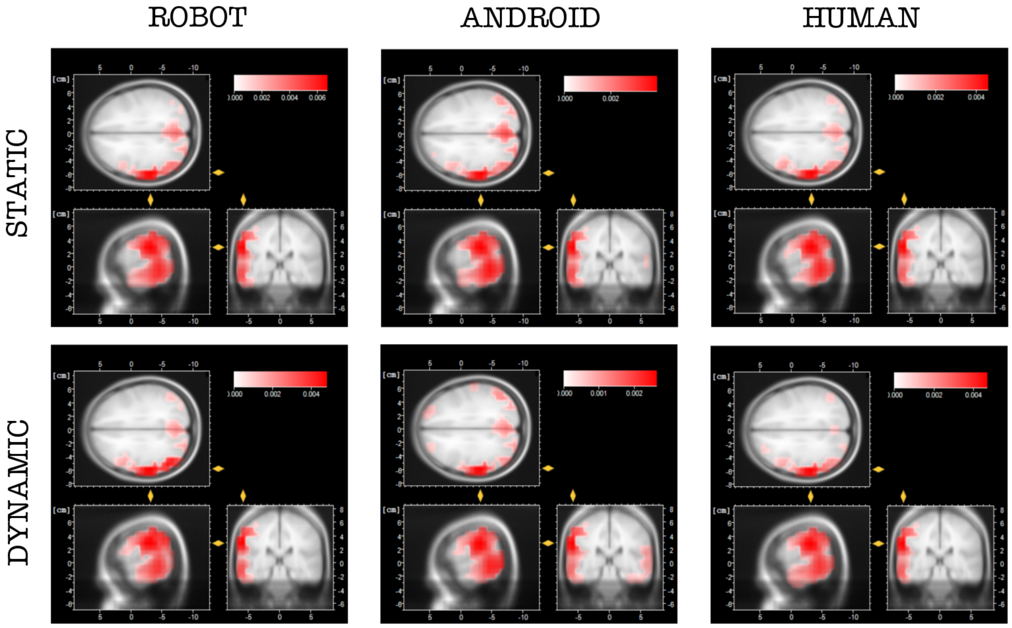

Figure 1.5: LORETA analysis in the N400 (370-600 ms) interval for each condition in the ERP experiment: The top row shows the source localization results in the static conditions of Robot, Android, and Human, and the bottom row shows the results in the dynamic conditions.

REFERENCES

Bartneck, C., Kanda, T., Ishiguro, H. and Hagita, N. (2009). My Robotic Doppelganger - A Critical Look at the Uncanny Valley Theory. *18th IEEE International Symposium on Robot and Human Interactive Communication*, RO-MAN2009, Toyama, IEEE.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision,* 10(4): 433-436.

Burleigh, T.J., Schoenherr, J.R. and Lacroix, G.L. (2013). Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in Human Behavior, 29*(3): 759-771.

Chen, H., Russell, R., Nakayama, K. and Livingstone, M. (2011). Crossing the 'uncanny valley': adaptation to cartoon faces can influence perception of human faces. *Perception, 39*(3): 378-386.

Cheetham, M., Suter, P. and Jancke, L. (2011). The human likeness dimension of the "Uncanny Valley Hypothesis": Behavioral and functional MRI findings. *Frontiers in Human Neuroscience, 5*: 126.

Cheetham, M. and Jancke, L. (2013). Perceptual and category processing of the Uncanny Valley hypothesis' dimension of human likeness: some methodological issues. *Journal of Visualized Experiments,* (76).

Cheetham, M., Pavlovic, I., Jordan, N., Suter, P. and Jancke, L. (2013). Category processing and the human likeness dimension of the Uncanny Valley Hypothesis: Eye-Tracking Data. *Frontiers in Psychology, 4*.

Delorme, A. and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1): 9-21.

Flach, L. M., Dill., V. and Lywkawka, C. (2012). *Evaluation of the uncanny valley in CG characters,* XI SB Games, Brazil.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*:127-138.

Ho, C. and MacDorman, K.F. (2008). Human emotion and the uncanny valley: a GLM, MDS, and Isomap analysis of robot video ratings. *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, Amsterdam, The Netherlands, ACM.

Ishiguro, H. (2006). Android science: conscious and subconscious recognition. *Connection Science, 18*(4): 319-332.

Kilner, J.M., Friston, K.J., Frith, C.D. (2007a). Predictive coding: an account of the mirror neuron system. *Cognitive Processing*, *8*:159-166.

Kilner, J.M., Friston, K.J. and Frith, C.D. (2007b). The mirror-neuron system: a Bayesian perspective. *Neuroreport*,*18*:619-623.

Kutas, M. and Federmeier, K.D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Reviews of Psychology*, 62: 621-647.

Lewkowicz, D. J. and Ghazanfar, A.A. (2012). The development of the uncanny valley in infants. *Developmental Psychobiology, 54*(2): 124-132.

Li, A.X., Florendo, M., Miller, L.E., Ishiguro, H., Saygin, A.P. (2015). Robot form and motion influences social attention. *10th ACM/IEEE International Conference on Human-Robot Interaction*, Portland, USA.

MacDorman, K.F. and Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies, 7*(3): 297-337.

MacDorman, K.F., Green, R.D., Ho, C. and Koch, C.T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior, 25*(3): 695-710.

Macdorman, K.F. and Entezari, S.O. (in press). Individual differences predict sensitivity to the uncanny valley. *Interraction Studies*.

Matsuda, Y. T., Okamoto, Y., Ida, M., Okanoya, K., Myowa-Yamakoshi, M. (2012). Infants prefer the faces of strangers or mothers to morphed faces: an uncanny valley between social novelty and familiarity. *Biologial Letters, 8*(5): 725-728.

Mitchell, W.J., Szerszen, K.A., Lu, A.S., Schermerhorn, P.W., Scheutz, M. and MacDorman, K.F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *Iperception, 2*(1): 10-12.

Moore, R.K. (2012). A Bayesian explanation of the 'Uncanny Valley' effect and related psychological phenomena. *Scientific Reports*, *2*: 864.

Mori, M. (1970). The uncanny valley. *Energy, 7*(4): 33-35.

Nie, J., Park, M., Marin, A.L., Sundar, S.S. (2012). Can you hold my hand? Physical warmth in human-robot interaction. *Human-Robot Interaction*. Boston, Massachusetts, USA.

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

Pascual-Marqui, R.D., Michel, C.M. and Lehmann, D. (1994). Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. International Journal of Psychophysiology, *18*(1), 49-65.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision, 10*(4): 437-442.

Piwek, L., McKay, L.S., and Pollick, F.E. (2014). Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion. *Cognition, 130*(3): 271-277.

Poliakoff, E., Beach, N., Best, R., Howard, T., and Gowen, E. (2013). Can looking at a hand make your skin crawl? Peering into the uncanny valley for hands. *Perception, 42*: 998-1000.

Pollick, F.E. (2010). In search of the uncanny valley. *UCMedia 2009, LNICST 40*, 69-78.

Rao, R.P. & Ballard, D.H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*, 79–87.

Rosenthal-von der Pütten, A.M. and Krämer, N.C. (2014). How design characteristics of robots determine evaluation and uncanny valley related responses. *Computers in Human Behavior, 36*: 422-439.

Saygin, A.P. and Cicekli, I. (2002). Pragmatics in human-computer conversations. *Journal of Pragmatics, 34*(3): 227-258.

Saygin, A. P., Chaminade, T., Urgen, B.A. and Ishiguro, H. (2011). Cognitive neuroscience and robotics: A mutually beneficial joining of forces. *Robotics: Systems and Science,* Los Angeles, CA.

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. *Social Cognitive Affective Neuroscience, 7*(4): 413-422.

Saygin, A.P. (2012). What can the Brain Tell us about Interactions with Artificial Agents and Vice Versa? *Workshop on Teleoperated Androids, 34th Annual Conference of the Cognitive Science Society*, Sapporo, Japan.

Seyama, J. and Nagayama, R. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and Virtual Environments, 16*: 337-351.

Steckenfinger, S.A. and Ghazanfar, A.A. (2009). Monkey visual behavior falls into the uncanny valley. *Proceedings of the National Academy of Sciences of the United States of America 106*(43): 18362-18366.

Thompson, J.C., Trafton, J.G. and McKnight, P. (2011). The perception of humanness from the movements of synthetic agents. *Perception, 40*: 695–705.

Tinwell, A., Nabi, D.A. and Chalton, J.P. (2013). Perception of psychopathy and the Uncanny Valley in virtual characters. *Computers in Human Behavior, 29*(4): 1617-1625.

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. *Frontiers in Neurorobotics, 7*: 19.

Walters, M.L., Syrdal, D.S., Dautenhahn, K., te Boekhorst, R. and Koay, K.L. (2008). Avoiding the uncanny valley: Robot apparence, personality, and consistency of behavior in an attention seeking home-scenario for a robot companion. *Autonomous Robots*, *24*(2), 159-178.

Yamada, Y., Kawabe, T. and Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the “uncanny valley” phenomenon. *Japanese Psychological Research*, *55*(1), 20-32.

Yamamoto, K., Tanaka, S., Kobayashi, H., Kozima, H. and Hashiya, K. (2009). A non-humanoid robot in the “uncanny valley”: Experimental analysis of the reaction to behavioral contingency in 2-3 year old children. *Plos One*, *4*(9).

CHAPTER 2

Human Cortical Oscillatory Activity during Perception of Human and Robot Actions

Abstract

The perception of others' actions supports important skills such as communication, intention understanding, and empathy. Are mechanisms of action processing in the human brain specifically tuned to process biological agents? Humanoid robots can perform recognizable actions, but can look and move differently from humans and as such, can be used in experiments to address such questions. Here, we recorded EEG as participants viewed actions performed by three agents. In the Human condition, the agent had biological appearance and motion. The other two conditions featured a state-of-the-art robot in two different appearances: Android, which had biological appearance but mechanical motion, and Robot, which had mechanical appearance and motion. We explored whether sensorimotor mu (8-13 Hz) and frontal theta (4-8 Hz) activity exhibited selectivity for biological entities, in particular for whether the visual appearance and/or the motion of the observed agent was biological. Sensorimotor mu suppression has been linked to the motor simulation aspect of action processing (and the human mirror neuron system, MNS), and frontal theta to semantic and memory-related aspects. For all three agents, action observation induced significant attenuation in the power of mu oscillations, with no difference between agents. Thus, mu suppression, considered an index of MNS activity, does not appear to be selective for biological agents. Observation of the Robot resulted in greater frontal theta activity compared to the Android and the Human, whereas the latter two did not differ from each other. Frontal

theta thus appears to be sensitive to visual appearance, suggesting agents that are not sufficiently biological in appearance may result in greater memory processing demands for the observer. Studies combining robotics and neuroscience such as this one can allow us to explore neural basis of action processing on the one hand, and inform the design of social robots on the other.

2.1 INTRODUCTION

From dolls and statues, to modern horror and science fiction stories, humans have long been preoccupied with creating other entities in their likeness. Advances in technology now allow us to create increasingly realistic and interactive humanoid agents. Lifelike humanoid robots are becoming commonplace, and assistive technologies based on social robotics are being developed for many application domains (e.g., Kanda et al., 2004; Coradeschi et al., 2006). Research on how humans perceive, respond to and interact with these agents is therefore increasingly important. However little is understood about human social cognition in this new, wider context. An interdisciplinary perspective on social robotics is needed, since this field will impact many areas of research, as well as issues of public concern in the near future, for example in domains such as education and healthcare (Billard et al., 2007; Dautenhahn, 2007; Mataric et al., 2009). Here, we provide hypotheses and data from cognitive and social neuroscience to study the perception of humanoid robots. Our goal is on the one hand to improve our understanding of human social cognition, and on the other, to help engineers and designers develop robots that are well-suited to their application domains.

2.1.1 Action Understanding and the Brain

Understanding the movements and actions of others is critical for survival in most species and for social cognition in primates. For humans, these processes are building blocks for important higher-order social skills, such as coordination, communication, intention understanding, and empathy (Blakemore & Decety, 2001; Iacoboni & Dapretto, 2006; Knoblich et al., 2006). A prominent idea regarding how the nervous system achieves the goal of "understanding others" is *motor simulation*. According to this theory, an action is understood by mapping the visual representation of an observed action to the observers' own motor representations (Rizzolatti et al., 2001). This view has become more widespread following the discovery of *mirror neurons (MNs)* in macaque premotor cortex (Di Pellegrino et al. 1992, Gallese et al. 1996, Rizzolatti et al. 1996). MNs are cells that fire both during the execution of an action, and during the observation of the same action performed by another agent, thereby providing a neural basis for motor resonance. For instance a mirror neuron that fires as the monkey cracks a peanut, can also fire as the monkey observes someone else crack a peanut. The neural network in the human brain supporting action and body movement processing is generally referred to as the *mirror neuron system (MNS)* – sometimes also as *action observation network* or *action perception system* – and corresponds to a set of areas in temporal, parietal and frontal cortices (Rizzolatti et al., 2001; Saygin, Wilson, Hagler, Bates, & Sereno, 2004; Grafton & Hamilton, 2007; Saygin, 2007; Cattaneo, Sandrini, & Schwarzbach, 2010; van Kemenade, Muggleton, Walsh & Saygin, 2012; Cook et al., in press). The MNS received considerable attention in the past two decades as a possible neural basis for action understanding, social cognition, empathy, and communication, and has been discussed in relation to disorders affecting social functions such as autism (Iacoboni & Dapretto, 2006).

Although the majority of studies on human MNS have involved functional magnetic resonance imaging (fMRI) as a method of investigation, there is also a body of evidence from multiple temporally-sensitive methodologies including motor-evoked potentials, magnetoencephalography (MEG), and electroencephalography (EEG) that motor system is involved during action observation (Fadiga et al., 1995; Hari et al., 1998; Cochin et al., 1999; Babiloni et al., 2002 Pineda, 2005; Hari, 2006; Orgs et al., 2008; Kilner et al., 2009, Perry & Bentin, 2009; Press et al., 2011). EEG studies in particular have revealed another index of human MNS activity known as *mu suppression*, which can be measured non-invasively via electroencephalogram (EEG), with electrodes placed on the scalp. Mu suppression refers to an attenuation in the power of the EEG in the alpha frequency range (8-13 Hz) measured over sensorimotor cortex and, like mirror neuron activity, is observed both during action execution and action observation (Cochin et al., 1999; Babiloni et al., 2002 Pineda, 2005; Hari, 2006; Orgs et al., 2008; Perry & Bentin, 2009). There is a growing body of literature that has revealed the functional properties of the sensorimotor mu suppression. Specifically, it has been suggested that mu suppression might have a role in social interactive contexts in addition to passive action observation (Tognoli et al., 2007; Dumas et al., 2012; Naeem et al., 2012; Silas et al., 2012), and that sub-bands of the mu rhythm might have different functional properties (Naeem et al., 2012). In an attempt to understand the relation between the mu suppression and the MNS, studies using both fMRI and EEG have argued that attenuations in the power of the EEG mu rhythm and fMRI activity in nodes of the MNS likely index the activity of the same underlying neural populations (Arnstein et al., 2011; Braadbaart et al., 2013), although it is worth noting mu suppression has also been correlated with brain areas other than the MNS (Mizuhara & Inui, 2011).

Although the 8-13 Hz oscillations have been the most implicated frequency band in EEG studies of action observation, a thorough understanding of the mechanisms of action observation and of the functional properties of this neural system can benefit from considering other dependent measures whose functional significance in cognition is well studied. As mentioned above, one of the most influential mechanistic explanation of action observation, the motor simulation framework, posits that we understand others' actions by mapping the visual input of the seen action to our own sensorimotor representations (Rizzolatti et al., 2001). For meaningful actions, during this mapping process, one also needs to activate the existing semantic representations of actions, and compare them with the current visual input and/or the representations evoked during motor simulation (Baressi & Moore 1996). If there is a match between the seen action's meaning and existing long-term memory representations, this can result in successful recognition of the action; if there is no match (e.g., in the case of actions or agents that have not been encountered before, and thus do not have a memory trace), the newly encountered item will need to be encoded into long-term memory. Thus, the entire process of action understanding requires the interplay of perceptual, motor, and memory processes.

Although memory is an essential part of action understanding (and the processing of meaningful stimuli in general), most studies to date have approached the issue implicitly (e.g., Umilta et al., 2001). However, both human behavioral and neuroscience studies (e.g., Carmo et al., 2012; Casile & Giese, 2006; Stefan et al., 2005) and robotics studies (e.g., Wermter & Elshaw, 2003; Ugur, Oztop, & Sahin, 2011) have highlighted a role for memory processes in action understanding, and there is growing interest in specifying the role of learning and memory in action perception and related brain systems (Cook et al., in press). EEG theta oscillations have been investigated in the context of memory processes, but have not been studied thoroughly in relation to action understanding. Given the crucial role of memory for action understanding within the motor simulation framework, we believe it is time to incorporate what we know about the functional significance of theta activity in studying action processing. Thus, in the current study, we also explored theta oscillations (4-8 Hz), which, especially at frontal sites, are thought to index memory encoding and retrieval in both linguistic and non-linguistic contexts (Hald et al., 2006; Osipova et al., 2006; Davidson & Indefrey, 2007; Bastiaansen et al., 2008; Shahin et al., 2009; Crespo-Garcia et al., 2010; Klimesch et al., 2010; Zion-Golumbic et al., 2010; Atienza et al., 2011). Specifically, theta activity has been reported to increase during encoding of information into long-term memory, and during retrieval of information from long-term memory (see review Klimesch et al., 2010). Zion-Golimbic et al. (2010) also reported that theta power increase reflects the utilization of information from long-term memory during processing of visual stimuli. Exploration of theta oscillations during action processing could be informative given the automatic employment of memory processing during action observation, and given that there is almost no work on theta oscillations in relation to action observation.

2.1.2 Cognitive Neuroscience and Robotics

The cognitive neuroscience of action perception, and especially the MNS, has received intense interest from neuroscientists in the last two decades, and we can now use the accumulated knowledge in this field to study how the human brain supports human-robot interaction. Conversely robotics can help research on the human brain by allowing us to test functional properties of the MNS and other brain areas that support action understanding.

One question that has been of interest since the identification of the MNS is whether the system is selectively tuned to process the actions of biological agents. For example, we may ask, during perception of or interactions with robots, does the brain rely on the same or distinct processes as with perception of or interactions with biological agents? The neuroscience-based theory of motor simulation argues that a visually perceived body movement or action is mapped onto the perceiving agent's sensorimotor neural representations, and "an action is understood when its observation causes the motor system of the observer to 'resonate" (Rizzolatti et al., 2001). But what are the boundary conditions for 'resonance'? What kinds of agents or actions lead to the simulation process? Is biological appearance important? Is biological motion? Artificial agents such as robots can be important experimental stimuli to test such hypotheses since robots can perform recognizable actions like biological agents, but can differ from biological agents in some other aspects (e.g., on how they appear or how they move – see below).

The neuroscience literature on the perception of robots has not revealed consistent results (Kilner et al., 2003; Chaminade & Hodgins, 2006; Gazzola et al., 2007; Oberman et al., 2007; Press et al., 2007; Chaminade et al., 2007). Some studies have reported that artificial agents' actions apparently affect the observers' own motor processing, or activity within the MNS, whereas others have argued that the MNS either does not respond, or responds weakly if the perceived actor is not human, including a clear claim that the MNS is only "mirror" for biological actions (Tai et al., 2004).

Conversely, neuroscience research on human observation of and interaction with robots can be invaluable to social robotics researchers since an important issue in the growing field of personal and social robotics is how to design robots that are likely to be socially accepted by their human companions. Research on the neural basis of social cognition using robots can provide valuable insights to advance the field of robot design and human-robot interaction by identifying the critical qualities that a robot should have, and eventually to guide the building of "neuroergonomic" robots that people are comfortable to interact with (Saygin et al., 2011).

2.1.3 Brain Activity and Robot Design

Here, we explored human brain activity evoked by humans and robots. Robots can have a range of appearance and movement patterns – but at the same time, they can be perceived as carrying out recognizable actions. Is biological appearance or biological movement necessary for engaging human brain systems that support social cognition? Does robot perception require additional memory processing demands? Robots can allow us to ask such questions and to test whether particular brain systems are selective for or sensitive to the presence of a human, or an agent with a humanlike form, or whether they respond similarly regardless of the agent performing the action.

Given that action observation is important for imitation learning and higher-level social skills, we hypothesized that human-likeness of the observed agent (i.e., the degree of similarity between the observer and the observed agent) could be important for the MNS. Indeed, motor resonance theory would predict increased humanlikeness would lead to more effective or efficient simulation (e.g., Buccino et al., 2004, Calvo-Merino et al., 2006; Casile et al., 2010). On the other hand, in artificial agents, human resemblance is not necessarily always a positive feature. The “uncanny valley” (UV) hypothesis suggests that as a robot is made more human-like, the reaction to it becomes more and more positive, until a point is reached at which the robot becomes oddly repulsive (Mori, 1970). This phenomenon is well known to roboticists and animators, but its scientific understanding remains incomplete, although there is a growing body of research on the topic, with some recent contributions from the behavioral and neural sciences (e.g., Cheetham, Suter, & Jancke, 2011; Ho, MacDorman, & Dwi Pramono, 2008; Lewkowicz & Ghazanfar, 2012; MacDorman & Ishiguro, 2006; Saygin et al., 2012; Steckenfinger & Ghazanfar, 2009; Thompson, Trafton, & McKnight, 2011; Tinwell, Grimshaw, Nabi, & Williams, 2011).

Most studies on the observation of robot actions had used very basic robot arms consisting of a stick/body and a claw, akin to rudimentary industrial robot arms, performing grasping or other simple movements. Therefore, the results are not sufficient to make conclusions regarding social humanoid robots that are being developed today. To overcome these limitations of previous work, we developed well-controlled stimuli based on state-of-the-art humanoid robots developed by an interdisciplinary team. Furthermore, our hypotheses, stimuli and experimental design focused on whether the seen agent had biological (human-like) appearance, whether the agent's body movements were biological, plus whether their appearance and movements matched (Saygin et al., 2012). We used human EEG cortical oscillatory activity in the alpha/mu and theta frequency bands as dependent measures in the present study. In addition to asking functional questions about action processing and social cognition, we also hope to shed new light onto the functional significance of these dependent measures in relation to action observation. For instance, are cortical theta and mu oscillations sensitive to the sensory properties of the stimuli, or to higher-level cognitive processes? In particular, we investigated whether cortical theta and mu oscillations are modulated by the humanlikeness of the observed agent. We characterized humanlikeness in two different ways: in terms of appearance and in terms of motion. Participants watched videos of three agents as their EEG was recorded: Human, Android, and Robot. Human had biological appearance and movement, Android had biological appearance and mechanical movement, and Robot had mechanical appearance and mechanical movement (see Figure 2.1, Methods, and Saygin et al., 2012 for more detail).

We hypothesized that if mu suppression is influenced by the specific visual properties of the seen action, we might find a difference between the actions of the different agents based on their differing appearance and/or motion characteristics. If on the other hand mu suppression reflects higher-level processes related to the meaning of the action, then the agents might not differ from each other since they all perform the same recognizable actions despite their different degrees of humanlikeness. For theta activity, we hypothesized that its power would be modulated by the humanlikeness of the observed agent, reflecting the processing demands of mapping the visual input into existing semantic representations. Since in the context of action processing, people are more familiar with human actors than robot actors, we hypothesized memory processes would differ depending on the agent’s appearance. More specifically, we hypothesized that the power of the theta oscillations would decrease as a function of the humanlikeness of the observed agent, since observation of relatively unfamiliar stimuli would result in greater memory processing demands (Hald et al., 2006; Zion-Golumbic et al., 2010; Atienza et al., 2011). We thus expected that observation of the Robot would result in increased theta activity compared to the Human, since the humanlike appearance of the agent would facilitate access to semantic representations related to human action. The Android condition, which features humanlike appearance but nonhuman motion, additionally allows us to ask whether or not the humanlikeness of the motion is a modulator of memory processes.

In sum, the aim of the study was three-fold. First, by manipulating various features of the observed agent, we aim to improve our understanding of the functional significance of EEG mu and theta oscillations during action observation and their relation to the MNS. Second, using robots as experimental stimuli in the presence of existing knowledge in cognitive neuroscience of action perception, we aim to inform robotics about how humans respond to robots of varying degrees of humanlikeness, what dependent measures could be used as gold-standards for social robotics research, and accordingly for guiding the design of robots in the long-term. Finally, the current study allows us to do cross-methodology comparison, as we previously reported an fMRI study utilizing the same agents as stimuli (Saygin et al., 2012).

2.2 MATERIAL AND METHODS

2.2.1 Participants

Twelve right-handed adults (3 females; mean age = 23.4; SD = 4.7) from the student community at the University of California, San Diego participated in the study. Participants had normal or corrected-to-normal vision and no history of neurological disorders. We recruited only those participants who had no experience working with robots in order to minimize possible effects of familiarity or expertise on our results (MacDorman, Vasudevan, & Ho, 2009). Informed consent was obtained in accordance with the UCSD Human Research Protections Program. Participants were paid $8 per hour or received course credit.

2.2.2 Stimuli

Stimuli were video clips of actions performed by the humanoid robot Repliee Q2 (in Robotic and Human-like appearance, Figure 2.1 right and middle images, respectively) and by the human 'master', after whom Repliee Q2 was modeled (Figure 2.1 left image). We refer to these agents as the Robot, the Android (dressed up robot), and the Human conditions (even though the former two are in fact the same robot).

Repliee Q2 has 42 degrees of freedom and can make face, head and upper body movements (Ishiguro, 2006). The robot's movements are mechanical or "robotic", and do not match the dynamics of biological motion. The same movements were videotaped in two appearance conditions. For the Robot condition, Repliee Q2's surface elements were removed to reveal its wiring, metal arms and joints, etc. The silicone 'skin' on the hands and face and some of the fine hair around the face could not be removed but was covered.

The movement kinematics for the Android and Robot conditions was identical, since these conditions comprised the same robot, carrying out the very same movements. For the Human condition, the female adult whose face was used in constructing Repliee Q2 was videotaped performing the same actions. All agents were videotaped in the same room with the same background. Video recordings were digitized, converted to grayscale and cropped to 400x400 pixels. Videos were clipped such that the motion of the agent began at the first frame of each 2-second video.

In summary, we had three agents and varied form and motion of the observed agent: A Human with biological appearance and motion, an Android with biological appearance and mechanical motion, and a Robot with mechanical appearance and motion. Due to the considerable technical difficulty in developing these stimuli and limitations inherent to the robot systems we worked with, we did not have a fourth condition (i.e. an agent with a well-matched mechanical appearance and biological motion) that would make our experimental design 2 (motion) x 2 (appearance).

2.2.3 Procedure

Before starting EEG recordings, participants were presented with all the action stimuli and were informed as to whether each agent was human or robot. Since prior knowledge can induce cognitive biases against artificial agents (Saygin & Cicekli, 2002), each participant was given exactly the same introduction to the study. Participants went through a short practice session before the experiment.

EEG was recorded as participants watched video clips of the three agents performing five different upper body actions (drinking from a cup, picking up and looking at an object, hand waving, introducing self, nudging). The experiment consisted of 15 blocks of 60 trials with equal number of videos of each agent and action (4 repetitions of each video in each block). Stimuli were presented in a pseudo-randomized order ensuring that a video was not repeated on two consecutive trials. Each participant experienced a different pseudo-randomized sequence of trials.

Stimuli were displayed on a 22' Samsung LCD monitor at 60 Hz using Python-based Vizard (Worldviz, Inc.) software. We displayed a gray screen with a fixation cross before the start of the video clip on each trial. Participants were instructed to fixate the blue fixation cross at the center of the screen for 700-1000 ms. Then the color of the fixation cross was changed to green and presented for 500-700 ms to inform participants of the upcoming video. A comprehension question was displayed every 6-10 trials after the video, asking participants a true/false question about the action in the just seen video (e.g. Drinking?). Since participants did not know whether they would receive a question during video presentation, this task allowed us to direct the subjects' attention to the stimuli, but not in a manner that might bias the results for any particular condition (behavioral performance in the task did not differ across conditions; all p values >0.1). Participants responded with a bimanual key press (Yes/No responses).

2.2.4 EEG Recording and Data Analysis

EEG was recorded at 512 Hz from 64 ActiveTwo Ag/AgCl electrodes (Biosemi, Inc.) following the International 10/20 system. The electrode-offset level was kept below 25 k-ohm. Four additional electrodes were placed above and below the right eye, and lateral to the eyes to monitor oculomotor activity. Two mastoid electrodes were placed behind the ears for re-referencing. The data were preprocessed with MATLAB and the EEGLAB toolbox (Delorme & Makeig, 2004). Each participant's data were first high-pass filtered at 1 Hz, low-pass filtered at 50 Hz, and re-referenced to average mastoids. Then the data were epoched ranging from 900 ms preceding video onset to 2000 ms after video onset, and were time-locked to the onset of the video clips. Atypical epochs of electromyographic activity were removed from further analysis by semi-automated epoch rejection procedures (kurtosis and probability-based procedures with standard deviation >/= 6). To remove eye-related artifacts, the data were decomposed by extended infomax ICA using the algorithm *binica*, and components that showed typical eye-related artifact characteristics were removed from the data. After preprocessing, data for each condition were transformed into a spectrographic image using 3-cycle Morlet wavelets in the 4-55 Hz frequency range at a number of frontal channels (F3 and F4), central channels (C3 and C4 over the sensorimotor cortex), and parietal channels (P3 and P4). The frontal and central channels were selected since these or neighboring electrodes were consistently reported in the literature on theta and mu oscillations, respectively (Oberman et al., 2007; Hald et al., 2006; Zion-Golumbic et al., 2010). For both mu and theta oscillations, these are the specific regions of interest that are related to our hypotheses regarding MNS and memory, and posterior electrodes for each frequency band are believed to have different functional significance. However, for completeness, we reported also on parietal channels to cover the posterior parts of the scalp. The mean power of the baseline period of the spectrographic images was removed from the power at each time point of the experimental trials.

2.2.5 Statistical Analysis

The spectral windows of mu and theta oscillations for statistical analyses were determined from the mean spectrographic images across all conditions in the 4-55 Hz frequency range and constrained by well-established windows of these cortical rhythms, which are 8-13 Hz for mu and 4-8 Hz for theta. The specific time windows for statistical analyses of the power of mu (8-13 Hz) and theta (4-8 Hz) oscillations were determined from the mean spectrographic image across all conditions, allowing us to test modulations in time periods of interest without introducing any bias for finding specific condition differences. For mu, mean alpha power in the time window of the mu attenuation (400-1400 ms. after stimulus onset) was extracted for each condition (Agent) and channel (C3: left hemisphere; C4: right hemisphere), and entered into a 3 (Agent) x 2 (Hemisphere) repeated measures ANOVA. For theta, the mean power in the time window of the theta increase (150-400 ms. after stimulus onset) was extracted for each condition (Agent) and channel (F3: left hemisphere; F4: right hemisphere) and entered into a 3 (Agent) x 2 (Hemisphere) repeated measures ANOVA. Although our hypotheses primarily related to the Agent manipulation (Robot, Android, Human), we also modeled Action (the five different actions) and Hemisphere (left, right) in ANOVA analyses to explore any modulation that may be specific to particular actions. These analyses are not reported since they did not reveal any action-specific effects or interactions involving Action, and the effects reported below for the 3x2 ANOVA did not change. Greenhouse-Geisser correction was applied to the ANOVAs whenever indicated. P-values reported below are two-tailed except for the comparisons of mu and alpha power against zero, where our hypotheses were one-tailed (i.e., we expected a decrease in mu power and an increase in theta power). Planned or posthoc t-test p-values were corrected for multiple comparisons.

In addition to our hypothesis-driven ANOVAs described above, for completeness, we also included ANOVAs for each of theta and mu oscillations in the other channel locations: (C3, C4) and (P3, P4) for theta; (F3, F4) and (P3, P4) for mu. Furthermore, given some recent experimental evidence that sub-bands of the mu oscillations might have different functional properties (Naeem et al., 2012), we ran two additional 3 (Agent) x 2 (Hemisphere) ANOVAs for lower (8-10 Hz) and upper (10-13 Hz) bands of the mu oscillations at channels C3 and C4.

2.2.6 Multivariate Pattern Analyses (MVPA)

In recent years, computational methods from machine learning have been used to analyze neuroimaging data as an alternative to conventional analyses (Kamitani & Tong, 2005; Haynes & Rees, 2006; Norman et al., 2006). The idea is to build a model (classifier) that can decode information recorded from the brain with neuroimaging. This is done by first training the model with a set of data labeled with class information (e.g., the conditions of the experiment) and allowing it to learn the patterns within the data, and then testing it with a separate set of data to see whether it can correctly predict unlabeled stimuli. Predictions with higher-than-chance accuracy indicate that there is sufficient information in the data that distinguishes the neural patterns corresponding to different conditions of an experiment. The advantage of these methods is that they are more sensitive to the differences between conditions since they consider the *patterns* of activity as the basic units of measurement, as opposed to an average of the activity, which may discard useful information. This is important in the context of the current study since there are discrepancies in the mu suppression literature, which might be due to the information lost by using the traditional analysis (i.e. averaging technique).

In order to explore subtle differences that may have been missed when analyzing mu and theta oscillations with traditional analyses as described above, we used MVPA using the pattern of mu activity and pattern of theta activity. We used support vector machines (Cortes & Vapnik, 1995) with a linear basis function and the LIBSVM software package (Chang & Lin, 2011) on mu oscillations at channels C3 and C4, and theta oscillations at channels F3 and F4 in three-way (Robot-Android-Human) and two-way classifications (Robot-Android, Robot-Human, Android-Human). The data that were fed into the classifier were time-frequency features in the frequency range 8-13 Hz and in the time interval 400-1400 ms for mu, and time-frequency features in the frequency range 4-8 Hz and in the time interval 150-400 ms for theta. The data were scaled before classification and 5-fold cross validation was applied in the classification procedure. The prediction accuracy (the number of correctly predicted trials) was used as the performance metric of the classifier. Each classification (Robot-Android-Human, Robot-Android, Robot-Human, Android-Human) was run three times for each subject and the average prediction accuracy of these three runs are reported. Above chance performance (corresponding to the 95% confidence interval) was 54.37% for the two-way classifications, and 37.59% for the three-way classification (Muller-Putz et al., 2008).

2.3 RESULTS

2.3.1 Mu Oscillations (8-13 Hz)

In our channels of interest, C3 and C4, action observation led to an increase in theta power shortly after stimulus onset (see Theta results below for quantified analyses), followed by an attenuation in alpha power starting around 400 ms, and becoming stronger around 600 ms after stimulus onset (Figure 2.2).

For observation of all agents' actions (Human as well as the two robot agents, Android and Robot), attenuation of the mu oscillations were robust and significant (Figure 2.3; C3: Human (Mean = -1.21, SD = 0.61), t(11) = -6.871, $p<0.001$; Android (Mean = -1.14, SD = 0.60), t(11) = -6.642, $p<0.001$; Robot (Mean = -1.21, SD = 0.74), t(11) = -5.675, $p<0.001$, and C4: Human (Mean = -1.09, SD = 0.71), t(11) = -5.328, $p<0.001$; Android (Mean = -1.15, SD = 0.65), t(11) = -6.11, $p<0.001$; Robot (Mean = -1.19, SD = 0.87), t(11) = -4.76, p = 0.001).

Suppression in alpha power was also observed in frontal and parietal channels over the scalp with greater suppression at parietal channels. Although, we report some results from other channels here for descriptive purposes, given the differential functional significance of frontal and posterior alpha, our focus will be on the hypothesis-driven analyses at channels C3 and C4.

2.3.1.1 ANOVA

Our primary comparison of interest was the 3 (Agent) x 2 (Hemisphere) repeated measures ANOVA at central channels C3 and C4, which revealed no main effect of Agent (F(2, 22) = 0.151) or Hemisphere (F(1,11) = 0.163) on the power of the mu oscillations (all p>0.1; Figure 2.3). There was no Agent x Hemisphere interaction (F(2,22) = 0.947, p>0.1).

When we explored lower (8-10 Hz) and upper (10-13 Hz) bands of the mu oscillations at the same channels separately, we again found no main effects or interactions (Lower Mu: Agent (F(1.376, 15.136) = 0.047), Hemisphere (F(1, 11) = 0.007), Agent x Hemisphere (F(2, 22) = 1.093); Upper Mu: Agent (F(2, 22) = 0.216), Hemisphere (F(1, 11) = 0.136), Agent x Hemisphere (F(2, 22) = 0.496) all p > 0.1). Similar 3 (Agent) x 2 (Hemisphere) repeated measures ANOVAs at frontal (F3, F4) and parietal channels (P3, P4) are reported here for completeness: There were no main effects or interactions (F3-F4: Agent (F(2, 22) = 0.210), Hemisphere (F(1, 11) = 0.110), Agent x Hemisphere (F(2, 22) = 1.334); P3-P4: Agent (F(2, 22) = 0.629), Hemisphere (F(1, 11) = 1.187), Agent x Hemisphere (F(2, 22) = 0.359), all p>0.1).

2.3.1.2 Multivariate Pattern Analysis (MVPA)

Multivariate pattern analyses of the mu suppression at channels C3 and C4 were performed to reveal any subtle modulations in alpha power over time that may have been missed due to averaging in the traditional analysis. For the three-way classification Robot-Android-Human (R-A-H), the average performance of MVPA for all subjects was not above chance (33.91% for C3 and 34.28% for C4). Pair-wise classifications Robot-Android (R-A), Robot-Human (R-H), and Android-Human (A-H) also resulted in chance-level performance on average (50.53%, 52.11%, and 49.77%, respectively for channel C3, and 50.95%, 51.31%, and 50.82%, respectively for channel C4).

2.3.2 Theta Oscillations (4-8 Hz)

In our channels of interests, at channels F3 and F4, action observation led to an increase in theta power starting at around 150 ms and lasting until about 400 ms after stimulus onset (Figure 2.4), followed by an attenuation in alpha power (see Mu results above for quantified analyses). For the Robot condition, the increase in theta was significant at both F3 and F4, (Figure 2.4; For F3**,** Mean = 0.71, SD = 1.05, t(11) = 2.322, $p<0.01$; for F4, Mean = 0.83, SD = 1.13, t(11) = 2.527, $p<0.01$). Observation of Android and Human actions also resulted in increased theta power that were either statistically significant or just at the cusp of significance (Figure 2.4; For F3, Human (Mean = 0.32, SD = 0.75), t(11) = 1.479, p = 0.054; Android (Mean = 0.45, SD = 0.88), t(11) = 1.774, p = 0.05; For F4, Human (Mean = 0.37, SD = 0.68), t(11) = 1.848, $p <0.05$; Android (Mean = 0.37, SD = 0.84), t(11) = 1.506, p = 0.053).

Increase in the power of theta oscillations was also observed at central and parietal channels over the scalp. Although, we reported results from all channels here, we based our discussion mainly on the hypothesis-driven results at channels F3 and F4 given the prior literature.

2.3.2.1 ANOVA

Our main comparison of interest, a 3 (Agent) x 2 (Hemisphere) repeated measures ANOVA at channels F3 and F4 revealed a significant main effect of Agent (F(1.350, 14.852) = 5.276), $p<0.05$, See Figure 2.5). Planned comparisons (paired t-tests) indicated theta oscillations were greater for the Robot condition compared with the Human (F(1,11) = 5.386, $p<0.05$) and the Android conditions (F(1,11) = 9.879, $p<0.01$). The effect of Hemisphere (F(1,11) = 1.144, p>0.1) or the Agent x Hemisphere interaction (F(1,11) = 3.196, p>0.1) were not significant.

Similar 3 (Agent) x 2 (Hemisphere) repeated measures ANOVAs at central and parietal channels are reported here for completeness: There was a main effect of Agent at central channels, but no effect of Hemisphere or interaction effect (C3-C4: Agent (F(1.133, 12.458) = 5.016), p<0.04, Hemisphere (F(1, 11) = 0.401), p>0.1, Agent x Hemisphere (F(2, 22) = 1.819)). The Agent effect reflected increased theta for the Robot, like similar to that found in frontal channels (see Figure 2.4). There were no main effects or interactions in parietal channels (P3-P4: Agent (F(1.260, 13.860) = 2.588), Hemisphere (F(1, 11) = 1.078), Agent x Hemisphere (F(2, 22) = 0.908), all p>0.1).

2.3.2.2 Multivariate Pattern Analysis (MVPA)

Although traditional analyses already revealed differences between agents, we applied multivariate pattern analyses on the theta oscillations at channels F3 and F4 for completeness. For the three-way classification Robot-Android-Human (R-A-H), the average performance of MVPA for all subjects was above chance (39.58% for C3 and 39.53% for C4). Pair-wise classifications Robot-Android (R-A) and Robot-Human (R-H) resulted in above chance performance on average (58.25% and 58.33%, respectively for channel F3, and 57.80% and 58.61%, respectively for channel F4). Android-Human (A-H) classification resulted in chance-performance on average (51.76% for channel F3 and 52.16% for channel F4). These MVPA results were thus in line with the results of the traditional analyses.

2.4 DISCUSSION

We investigated how the sensorimotor EEG mu rhythm that is considered to index human MNS activity, and the frontal theta activity that is implicated in memory processes are modulated by the humanlikeness of the agent being observed. Participants viewed three agents, a Human, and a state-of-the-art robot in two different appearances (as an Android and a Robot) performing the same recognizable actions. The Human had biological motion and appearance, whereas the Android had biological appearance and mechanical motion, and the Robot had mechanical motion and mechanical appearance (Figure 2.1). We hypothesized that any modulations of the oscillations by sensory features of the stimuli would be revealed as significant differences between the experimental conditions, based on the seen agents' differing appearance and motion characteristics. Specifically if these dependent measures are sensitive to the movement kinematics of the seen actor, then we would expect the Human condition to be distinguished from the others. If they are sensitive to the appearance, then Robot would be distinguished from the other agents, or there will be a degree of activity that corresponds to the degree of humanlikeness of the appearance of the agents. If they are sensitive to the congruence of the movement dynamics and appearance, then Android would be distinguished from the other agents since this condition features a humanlike appearance along with nonhuman motion dynamics, whereas the other agents feature congruent appearance and motion (both biological, or both mechanical). If on the other hand these dependent measures reflect higher-level processing related to the meaning of the actions and are not sensitive to the visual properties of the stimuli, then the agents might not differ from each other since they all perform the very same actions.

2.4.1 Mu Oscillations

We showed that the observation of the human agent as well as both of the robot agents resulted in robust and significant attenuations in the power of mu oscillations over the frequently reported sensorimotor areas. The magnitude of the attenuations was equivalent for all agents. This replicates and extends a previous mu suppression study that had used a simple robot hand (Oberman et al., 2007). Consistent with previous work on action observation, we did not find any hemispheric differences (Babiloni et al., 2012). Overall, our results show that the human MNS is unlikely to be selective only for other humans, since a commonly accepted measure of human MNS activity (EEG mu suppression) showed robust and significant modulations also when observing robot actions. These data also suggest that mu suppression might not be sensitive to early sensory stages of action processing, since the agents' differences in terms of their visual appearance and movement kinematics did not differentially affect mu power. Frontal and parietal sites also showed the same pattern of results as the sensorimotor channels, although it must be noted that alpha oscillations at these latter sites are not specifically linked to the MNS or action processing.

After exploring mu suppression with traditional statistical analyses adopted from previous work (e.g., Oberman et al., 2007), we also explored the data using machine learning and multivariate pattern analyses. The pattern activity has more information than the average activity (over time and frequency band) used in traditional analyses so more subtle differences can be picked up with a more sensitive approach (see Kamitani & Tong, 2005; Pereira et al., 2009; Norman et al., 2006 for discussion of such issues). Our primary goal in applying pattern analysis on mu oscillations was to reduce concerns readers may possibly have about lack of a difference between conditions being due to an insensitive analysis method. In other words, we wanted to pre-answer the question a reader may have about whether there could be subtler differences when the entire pattern gets taken into account, especially given that recent literature on the mu suppression field has started to include finer modulations (Naeem et al., 2012). The fact that we did not find differences in the “patterns” of mu suppression with this much more sensitive analysis method provides strong evidence that mu suppression is also found for observing the actions of humanoid robots. Mu suppression patterns do not appear to be sensitive to the early sensory stages of action processing (as evidenced by chance level performance for the Robot-Android-Human classification), in particular to the appearance (as evidenced by chance level performance for the Robot-Android classification) or the movement kinematics (as evidenced by chance-level performance for the Android-Human classification) of the observed agent.

To be clear, there may be other systems in the brain that are modulated by sensory properties of the seen stimuli, or even those that are selective for processing biological agents. Indeed, in related work (and in the theta results here), we have reported perceptual and neural processes that are sensitive to the properties of the seen action such as humanlike appearance or motion (e.g., Urgen et al., 2012; Saygin & Stadler, 2012). The mu suppression results here indicate however that the human MNS does not appear to respond differentially to the actions of humanoid robots and humans. Although a PET study had claimed the human MNS is “mirror only for biological actions” (Tai et al., 2004), several recent fMRI studies are consistent instead with our present results, and have reported that human MNS also responds to robot actions (e.g., Gazzola et al., 2007;

Saygin et al., 2012; Cross et al., 2012). In particular, Saygin et al. (2012), using very similar stimuli to the present study, found no difference between human and robot actions in premotor cortex, but showed that parietal cortex is sensitive to the congruence of the motion and appearance of the agent (as evidenced by significant differences in response to the Android). More broadly, these data are consistent with the view that the premotor cortex is largely insensitive to the surface properties of the stimuli depicting actions, but instead is more involved in computing goals and intentions (Rizzolatti et al., 2001; Cattaneo et al., 2010; Grafton & Hamilton, 2007). Human fMRI studies indicate that human premotor cortex responds to a wide range of action stimuli, including impoverished or simplified displays such as point-light biological motion or simple avatars (Saygin et al., 2004; Pelphrey et al., 2003). Since the mu rhythm appears to be insensitive to the visual aspects of the actions (i.e., the humanlike appearance and movement kinematics in the current study), cognitive and affective manipulations during passive action observation or social interactive contexts as evidenced by recent literature (Tognoli et al., 2007; Dumas et al., 2012; Naeem et al., 2012; Silas et al., 2012) would be more appropriate for future studies to better understand the functional properties of the mu rhythm. The fact that we did not find any difference between the different sub-bands of the mu rhythm further suggests that social interactive contexts may be suitable to study the functional properties of the mu rhythm (Naeem et al., 2012).

2.4.2 Theta Oscillations

For the frontal theta oscillations, we expected our meaningful stimuli to lead to increases in power, reflecting memory-related processing (i.e., accessing long-term memory representations to process the higher-level meaning of the action stimuli). In particular, we hypothesized that the power would decrease as a function of the humanlikeness of the observed agent, since observation of relatively unfamiliar agents could result in greater memory processing demands (Hald et al., 2006; Zion-Golumbic et al., 2010; Atienza et al., 2011). More specifically, we hypothesized that observation of the Robot would result in greater theta activity compared to the Human, as we expected the humanlike appearance of the agent would facilitate access to semantic representations related to human action. However it was also possible biological motion would also influence responses, in which case we would expect the Android condition to also differ from the Human.

Our analysis of the frontal theta activity indeed showed that observation of the Robot resulted in a significantly stronger increase in the power of theta oscillations (4-8 Hz) compared to the agents with humanlike appearance over frontal sites; the Human and Android did not differ from each other. Multivariate pattern analysis of the theta oscillations corroborated these results. Since the Robot was distinguished from the other agents by its nonhuman appearance, these results suggest that frontal theta activity is modulated by the appearance of the agent being observed during action observation. Our exploratory analysis in central sites revealed a similar pattern of results with the frontal sites; there were no agent differences over parietal sites.

Since theta oscillations reflect memory processes such as retrieval from long-term memory and encoding into long-term memory (see review Kahana et al., 2001, Klimesch et al., 2010), our results suggest that processing of the Robot resulted in greater demands on memory systems as compared to the other agents. This difference is best viewed as reflecting the interplay of perceptual processing and long-term memory, particularly during retrieval of items from semantic memory. A robotic appearance, especially in the context of actions that are typical for humans, is less likely to be associated with strong semantic links that can aid in the mapping of the visual input onto existing representations from long-term memory. The difficulty of integrating the visual input with existing semantic knowledge could manifest itself as increased frontal theta activity in comparison to the conditions with humanlike appearance. For the human stimuli, linking the visual input with semantic representations of human actions is likely to be less effortful, since participants have had existing semantic representations about actions developed over time by seeing other humans. This interpretation is consistent with previous work, which has found increased theta activity during the retrieval of semantic information from long-term memory, and especially sensitivity to semantic congruence in linguistic and non-linguistic contexts (Hald et al., 2006; Davidson & Indefrey, 2007; Bastiaansen et al., 2008; Shahin et al., 2009; Zion-Golimbic et al., 2010; Atienza et al., 2011; Steele et al., 2013). The similarity of the results for the Android with that of the Human suggests that the very humanlike appearance of the Android may have facilitated the activation of semantic representations about human actions, even though the motion of this agent was not humanlike (and was in fact the same as that for Robot), and even though participants knew that this agent was not a real human.

In a recent event-related potential (ERP) study (Urgen et al., 2012), we averaged the EEG time-locked to the onset of actions for the Human, Android and Robot conditions. While all action stimuli evoked a significant negativity called the N300/N400 component complex beginning at around 250 ms after stimulus onset over frontal channels, the amplitude of this component differed significantly for the Robot condition compared to the other agents, a parallel to the present results. Given the timing of the theta oscillations observed here, and the known function of these ERP components for semantic processing (Sitnikova et al., 2008; Wu and Coulson, 2011; Kutas and Federmeier, 2011), we conclude that a humanlike appearance facilitates (or a nonhuman appearance impedes) access to long-term memory representations related to action. The link between frontal theta and ERP components related to memory processes should be explored in future work. Furthermore, the addition of a fourth condition which presents a biological motion and mechanical appearance combination (which is missing in the current study due to technical difficulties) in addition to the conditions in the current study can be studied in future research to better understand the interaction between the appearance and motion parameters.

2.4.3 Cross-methods Comparison: EEG and fMRI with Human and Robot Action Stimuli

The present study allows us to compare our EEG time-frequency results to our previous fMRI work with a similar stimulus set (Saygin et al., 2012). The main finding of our fMRI study was that parietal regions of the human cortex (specifically bilateral anterior intraparietal sulcus which is part of the human MNS) responded significantly more to the Android agent, therefore to the congruence of form and motion of the agent being observed. Premotor regions of the MNS did not show selectivity neither for the form nor for the motion of the agents. Although EEG mu activity has been found to correlate with fMRI activity both in premotor cortex and parietal cortex (Arnstein et al., 2011), the present results, along with our fMRI work (Saygin et al., 2012) suggest that the mu rhythm might share more functional properties with the activity of premotor cortex than parietal cortex.

In the current study theta oscillations distinguished the Robot agent around 150-400 ms after stimulus onset. Although there was a region in left lateral temporal cortex (the extrastriate body area) that responded significantly less to the Robot agent in the fMRI data, based on the functional properties of this region, this activation is more likely to reflect visual stimulus properties rather than the memory-related processing indexed by the theta oscillations in the present study, or by event related potentials (Urgen et al., 2012). It is likely that EEG, with its milliseconds time resolution, can reveal effects that do not emerge in fMRI studies due to the limited time resolution of this latter method.

2.4.4 Human Qualities of Social Robots

Neuroscience research on human observation of and interaction with robots not only improves our understanding of the neural basis of social cognition but it can be invaluable to social robotics researchers. One important issue in the growing field of personal and social robotics is how to design robots that are likely to be socially accepted by their human companions. Broadly, there are two important design issues. The first is the visual properties of robots, i.e. how they should look on the surface. The second is the functional properties, i.e. how they should perform the functions they are designed to perform. Not only should we consider both visual properties and functional properties, but the combination (or interaction) of them might also be important for determining the eventual social acceptability of a robot. Therefore, research efforts for robotics from social, behavioral and neural sciences should focus on both of these aspects. In the present study, we kept the functioning of the robots constant (i.e. both robots successfully perform various goal-directed, recognizable human actions) and manipulated the visual properties. Our study, together with existing neuroimaging evidence, provides some insight to robotics about the visual human qualities of robots that will be designed to interact with humans. It seems that as long as the robot performs the action successfully and is of a sufficiently humanoid design to do so, it will likely be processed in a similar way in the MNS as other people (see also, Gazzola et al, 2007; Oberman et al., 2007; Saygin et al., 2012). Although mu suppression has been linked to the human MNS, it and MNS activity in general do not appear be the right dependent measure for comparing the visual properties of robots with human standards. We suggest that neuroscience studies with mu oscillations as dependent measure might consider using cognitive and affective manipulations to study robot perception or human-robot interaction. It is possible that mu oscillations can inform design issues about the functional properties of robots, rather than visual properties.

Analysis of frontal theta activity on the other hand suggests that theta oscillations can be used as a dependent measure to investigate responses to visual properties of artificial agents, in particular on the interplay between perceptual and memory processes. Our results showed that if the artificial agent possesses sufficient amount of human-likeness in terms of its appearance (Android vs. Robot), it is more likely to facilitate access to semantic representations pertaining to the seen stimuli (here, actions). If the seen agent is rather different from a human in terms of its visual appearance (as in the case of the Robot), it will result in greater processing demands in the observer. Our results suggest that movement kinematics might not be as important as the appearance in influencing the mapping process of the visual input to the existing long-term memory representations.

In general, future social, behavioral and neuroscience research on the perception of robots should distinguish the two dimensions, i.e. visual properties and functional properties (and an interaction of the two) when studying the social acceptability of robots. This would result in a more systematic study of the design issues about social robots and enable determination of the right dependent measures to be used as gold standards in human-robot interaction and robot design. This research in turn will inform social and cognitive neuroscience about the neural basis of the social skills humans have. Our study demonstrates that this interdisciplinary work is useful and fruitful, and progress in this direction will improve our understanding in both fields.

ACKNOWLEDGMENTS

Chapter 2, in full, is a re-print of the material as it appears in Frontiers in Neurorobotics 2013. Urgen, B.A., Plank, M., Ishiguro, H., Poizner, H., & Saygin, A.P. The dissertation author was the primary investigator and author of this paper. This research was supported by the Qualcomm Institute (formerly California Institute of Telecommunications and Information Technology (Calit2), Strategic Research Opportunities Award to APS, fellowship for BAU), Kavli Institute for Brain and Mind (Innovative Research Award to APS), NSF (CAREER Award BCS-1151805 to APS, and SBE-0542013 to Temporal Dynamics of Learning Center), DARPA (APS), and ONR (MURI Award N00014-10-1-0072 to HP). We thank co-authors Markus Plank, Hiroshi

Ishiguro, Howard Poizner and Ayse P. Saygin. We also thank Arthur Vigil and Joe Snider for assistance with the experimental setup, Intelligent Robotics Laboratory at Osaka University for help in the preparation of the stimuli, and Alvin Li, Wayne Khoe, Marta Kutas, Seana Coulson, Jamie Pineda, Chris Berka, and Scott Makeig for helpful discussion and feedback.

Figure 2.1: Still frames from the videos used in the experiment depicting the three actors: Human, Android, and Robot.

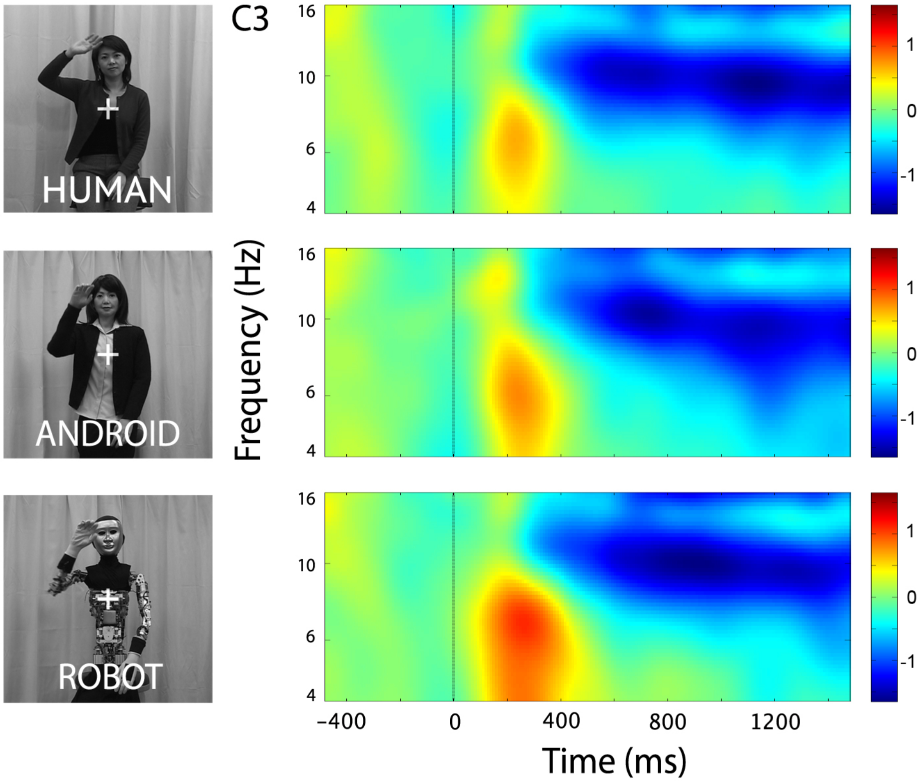

Figure 2.2: Time-frequency plots for the three conditions (Human, Android, Robot) at channel C3 (left) over somatosensory cortex (collapsed across actions). Plots for the right hemisphere (C4) were very similar and are not shown. The frequency-axis is log-spaced. The zero point on the time-axis indicates the onset of the action movies. Scaling unit is dB. Shortly after the onset of the action videos, we observed an increase in the theta frequency band (see also Figure 2.4), followed by an attenuation in the alpha frequency band that started around 400 ms, and grew stronger around 600 ms.

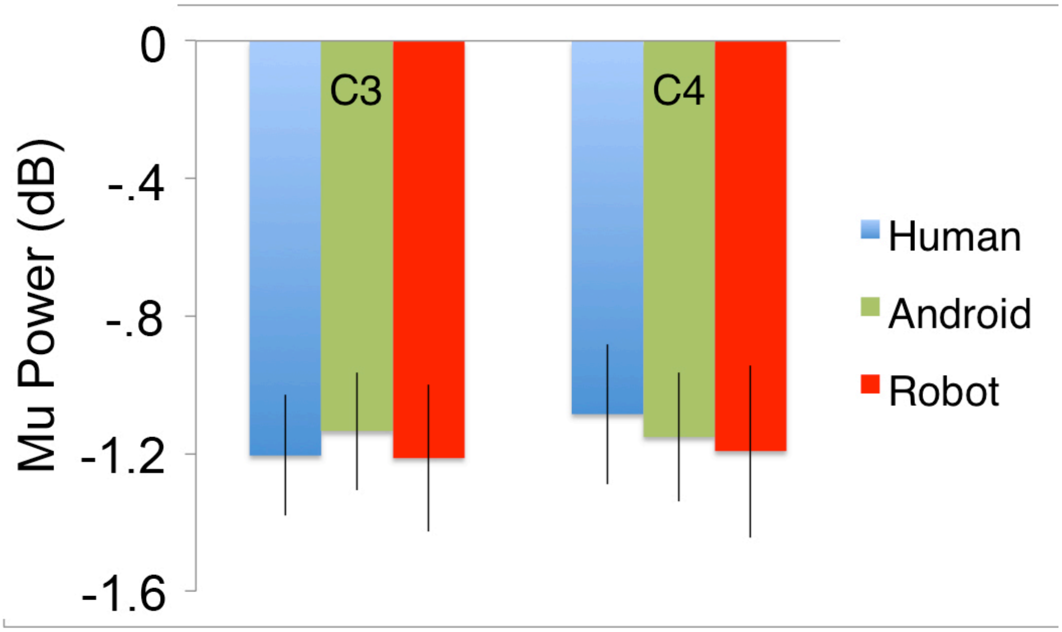

Figure 2.3: Attenuation in the power (in dB unit) of the mu (8-13 Hz) oscillations for the three conditions (Human, Android, Robot) plotted at each channel C3 and C4 over left and right somatosensory cortex, respectively. Error bars indicate the standard error of the mean. For both C3 and C4, all conditions led to statistically significant attenuation in mu power (all p's<0.05, see Results). There were no significant differences between agents (Human, Android, Robot) or hemispheres (C3, C4).

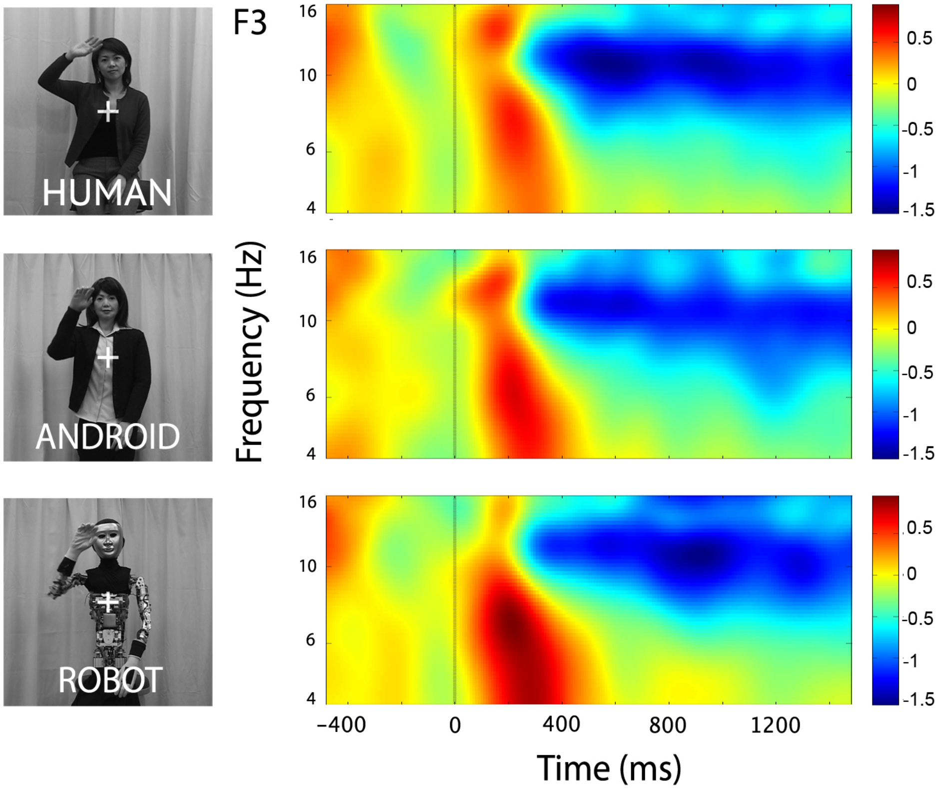

Figure 2.4: Time-frequency plots for the three conditions (Human, Android, Robot) at channel F3 (left) over frontal cortex (collapsed across actions). Plots for the right hemisphere (F4) were very similar and are not shown. The frequency-axis is log-spaced. The zero point on the time-axis indicates the onset of the action movies. Scaling unit is dB. Shortly after the onset of the stimuli, there was an increase in theta power, followed by a reduction in alpha power (see also Figure 2.2).

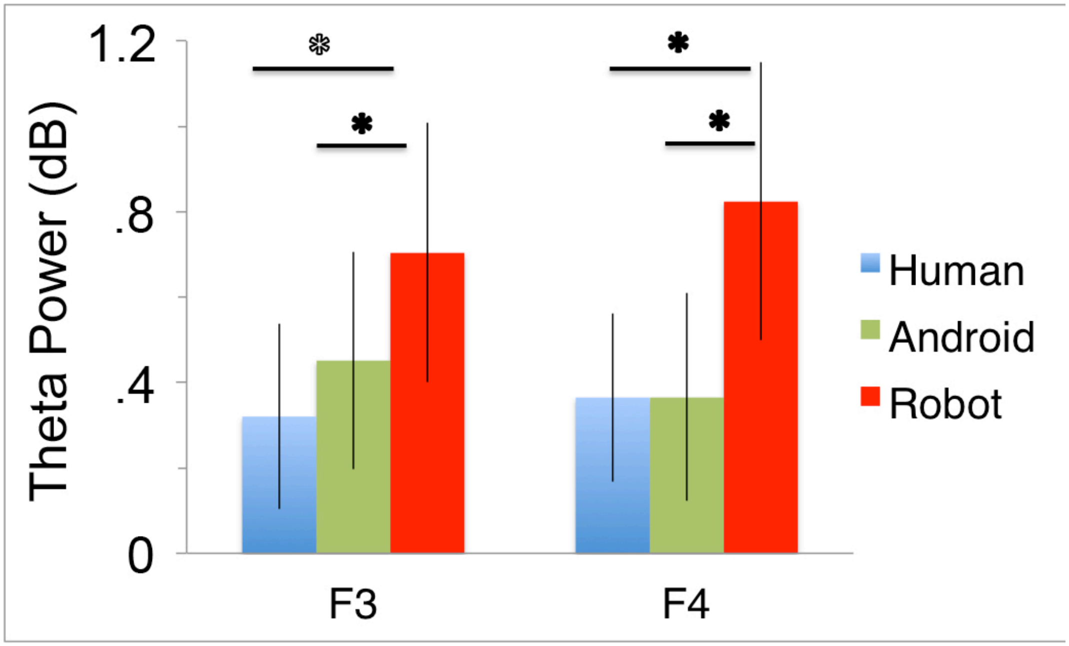

Figure 2.5: Power in the theta frequency range (4-8 Hz, in dB unit) for the three conditions (Human, Android, Robot) plotted at channels F3 and F4 (over left and right frontal regions, respectively). Error bars indicate the standard error of the mean. All conditions led to significant increase in theta power (all p's≤ 0.05, see Results). The Robot condition led to significantly increased theta power in comparison to Android conditions (F3: p<0.05; F4: p<0.05) and Human (F3: p=0.059; F4: p<0.05) (* indicates p<0.05 except the comparison marked in light gray * p = 0.059).

REFERENCES

Arnstein, D., Cui, F., Keysers, C., Maurits, N.M. and Gazzola, V. (2011). Mu suppression during action observation and execution correlates with BOLD in dorsal premotor, inferior parietal, and SI cortices. *Journal of Neuroscience, 31 (40)*, 14243–14249.

Atienza, M., Crespo-Garcia, M., & Cantero, J. L. (2011). Semantic congruence enhances memory of episodic associations: role of theta oscillations. *Journal of Cognitive Neuroscience*, *23*(1), 75–90.

Babiloni C, Babiloni F, Carducci F, Cincotti F, Cocozza G, Del Percio C, Moretti DV, Rossini PM (2002). Human cortical electroencephalography (EEG) rhythms during the observation of simple aimless movements: a high-resolution EEG study. *Neuroimage, 17 (2),* 559-572.

Barresi, J. and Moore, C. (1996). Intentional relations and social understanding. *Behavioral and Brain Sciences, 19*, 107-154.

Bastiaansen, M. C. M., Oostenveld, R., Jensen, O., & Hagoort, P. (2008). I see what you mean: theta power increases are involved in the retrieval of lexical semantic information. *Brain and Language*, *106*(1), 15–28.

Billard, A., Robins, B. Nadel, J. Dautenhahn, K. (2007). Building Robota, a mini-humanoid robot for the rehabilitation of children with autism. *Assist Technol. 19*, 37-49.

Blakemore, S.J., and Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience, 2*:561–66.

Braadbaart, L., Williams, J. H., & Waiter, G. D. (2013). Do mirror neuron areas mediate mu rhythm suppression during imitation and action observation? *International Journal of Psychophysiology*. *89*, 99-105.

Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., et al. (2004). Neural circuits involved in the recognition of actions performed by nonconspecifics: an FMRI study. *Journal of Cognitive Neuroscience, 16(1),* 114-126.

Calvo-Merino, B., Grezes, J., Glaser, D.E., Passingham, R.E., Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Current Biology, (16),* 1905-1910.

Carmo, J. C., Rumiati, R. I., & Vallesi, A. (2012). Understanding and imitating unfamiliar actions: distinct underlying mechanisms. *PloS One*, *7*(10).

Casile, A. and Giese, M.A. (2006). Nonvisual motor training influences biological motion perception. *Current Biology, 16*, 69-74.

Casile, A., Dayan, E., Caggiano, V., Hendler, T., Flash, T. & Giese, M. A. (2010). Neuronal encoding of human kinematic invariants during action observation. *Cerebral Cortex, 20 (7)*, 1647-1655.

Cattaneo, L., Sandrini, M., and Schwarzbach, J. (2010). State-Dependent TMS reveals a hierarchical representation of observed acts in the temporal, parietal, and premotor cortices. *Cerebral Cortex, 20(9),* 2252-2258.

Chaminade, T., & Hodgins, J. K. (2006). Artificial agents in social cognitive sciences. *Interaction Studies, 7*(3), 347-353.

Chaminade, T., Hodgins, J., & Kawato, M. (2007). Anthropomorphism influences perception of computer-animated characters' actions. *Social Cognitive and Affective Neuroscience, 2*(3), 206-216.

Chang, C. and Lin, C. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1--27:27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Cheetham M., Suter P., Jancke L. (2011). The human likeness dimension of the "uncanny valley hypothesis": behavioral and functional MRI findings. *Frontiers in Human Neuroscience*, 5:126.

Cochin, S., Barthelemy, C., Roux, S., & Martineau, J. (1999). Observation and execution of movement: similarities demonstrated by quantified electroencephalography. *European Journal of Neuroscience, 11(5)*, 1839-1842.

Cook, R., Bird, G., Catmur, C., Press, C., & Heyes, C. (in press). Mirror Neurons : From origin to function. *Behavioral and Brain Sciences*.

Coradeschi, S. Ishiguro, H., Asada, M., Shapiro, S.C., Thielscher, M., Breazeal, C., Mataric, M.J. and Ishida, H. (2006). Human-inspired robots. *IEEE Intelligent Sys. 21*, 74-85.

Cortes, C. and Vapnik, V.N. (1995). Support-vector networks. *Machine Learning*, 20.

Crespo-Garcia, M., Cantero, J. L., Pomyalov, a, Boccaletti, S., & Atienza, M. (2010). Functional neural networks underlying semantic encoding of associative memories. *NeuroImage*, *50*(3), 1258–70.

Cross, E., Liepelt, R., Hamilton, A., Parkinson, J., Remsey, R., Stadler, W. & Prinz, W. (2012). Robotic movement preferentially engages the action observation network. *Human Brain Mapping, 33 (9)*, 2238-2254.

Dautenhahn, K. (2007). Socially intelligent robots: dimensions of human-robot interaction. *Philos. Trans. R. Soc. Lond. B Biol. Sci., 362*, 679-704.

Davidson, D. J., & Indefrey, P. (2007). An inverse relation between event-related and time-frequency violation responses in sentence processing. *Brain Research*, *1158*, 81–92.

Delorme, A. & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134(1)*, 9-21.

Di Pellegrino G, Fadiga L, Fogassi L, Gallese V, Rizzolatti G. 1992. Understanding motor events: a neurophysiological study. *Experimental Brain Research, 91*:176–80.

Dumas, G., Martinerie, J., Soussignan, R., & Nadel, J. (2012). Does the brain know who is at the origin of what in an imitative interaction?. *Frontiers in Human Neuroscience*, 6:128.

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: a magnetic stimulation study. *Journal of Neurophysiology, 73(6)*, 2608-2611.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119*, 593-609.

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

Grafton, S. T., & Hamilton, A. F. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Human Movement Science, 26(4),* 590-616.

Hald, L.A., Bastiaansen, M.C.M., & Hagoort, P. (2006). EEG theta and gamma responses to semantic violations in online sentence processing. *Brain and Language*, *96*(1), 90–105.

Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Salenius, S., & Rizzolatti, G. (1998). Activation of human primary motor cortex during action observation: a neuromagnetic study. *Proceeding of National Academy of Sciences, 95(25)*, 15061-15065.

Hari, R. (2006). Action–perception connection and the cortical mu rhythm. *Progress in Brain Research, 159*, 253-260.

Haynes, J. D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience, 7(7)*, 523-534.

Ho, C.-C., MacDorman, K. F., & Dwi Pramono, Z. A. D. (2008). Human emotion and the uncanny valley: a GLM, MDS, and Isomap analysis of robot video ratings. *Proceedings of the 3rd ACM/IEEE international conference on human-robot interaction*. Amsterdam, The Netherlands: ACM.

Iacoboni, M. & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience, 7(12),* 942-951.

Ishiguro, H. (2006). Android science: conscious and subconscious recognition. *Connection Science, 18(4)*, 319-332.

Kahana, M. J., Seelig, D., & Madsen, J. R. (2001). Theta returns. *Current Opinion in Neurobiology*, *11*(6), 739–44.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience, 8(5),* 679-685.

Kanda, T., Ishiguro, H., Imai, M. and Ono, T. (2004). Development and evaluation of interactive humanoid robots. *Proc. IEEE, 92,* 1839-1850.

Kilner, J. M., Paulignan, Y., & Blakemore, S. J. (2003). An interference effect of observed biological movement on action. *Current Biology 13*(6), 522-525.

Kilner, J. M., Marchant, J. L., & Frith, C. D. (2009). Relationship between activity in human primary motor cortex during action observation and the mirror neuron system. *PLoS ONE, 4(3)*, e4925.

Klimesch, W., Freunberger, R., & Sauseng, P. (2010). Oscillatory mechanisms of process binding in memory. *Neuroscience and Biobehavioral Reviews*, *34*(7), 1002–14.

Knoblich, G., Thornton, I.M., Grosjean, M., Shiffrar, M., eds. (2006). *Human Body Perception from the Inside Out.* New York: Oxford Univ. Press

Kutas, M. and Federmeier, K.D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology, 62*, 621-647.

Lewkowicz D. J. and Ghazanfar A. A. (2012). The development of the uncanny valley in infants. *Developmental Psychobiology, 54*, 124–132.

MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies, 7*(3), 297-337.

MacDorman, K. F., Vasudevan, S. K., & Ho, C.-C. (2009). Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures. *AI & Society, 23(4),* 485-510.

Mataric, M., Tapus, A., Winstein, C. and Eriksson, J. (2009). Socially assistive robotics for stroke and mild TBI rehabilitation. *Stud. Health. Technol. Inform, 145*, 249-262.

Mizuhara, H. and Inui, T. (2011). Is mu rhythm an index of the human mirror neuron system? A study of simultaneous fMRI and EEG. In R. Wang, F. Gu (eds), *Advances in Cognitive Neurodynamics (II)* (123-127). Springer Netherlands.

Mori, M. (1970). The uncanny valley. *Energy, 7*(4), 33-35.

Muller-Putz, G.R., Scherer, R., Brunner, C. Leeb, R., and Pfurtscheller, G (2008). Better than random: A closer look on BCI results. *International Journal of Biomagnetism, 10(1)*, 52-55.

Naeem, M., Prasad, G., Watson, D. R., & Kelso, J. A. (2012). Electrophysiological signatures of intentional social coordination in the 10–12 Hz range. *Neuroimage, 59(2)*, 1795-1803.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences, 10(9),* 424-430.

Oberman, L. M., McCleery, J. P., Ramachandran, V. S. & Pineda, J. A. (2007). EEG evidence for mirror neuron activity during the observation of human and robot actions: Toward an analysis of the human qualities of interactive robots. *Neurocomputing, 70,* 2194-2203.

Orgs, G., Dombrowski,J., Heil, M. and Jansen-Osmann, P. (2008). Expertise in dance modulates alpha/beta event-related desynchronization during action observation. *European Journal of Neuroscience. 27(2)*, 3380-3384.

Osipova, D., Takashima, A., Oostenveld, R., Fernández, G., Maris, E., & Jensen, O. (2006). Theta and gamma oscillations predict encoding and retrieval of declarative memory. *Journal of Neuroscience*, 26(28), 7523–31.

Pelphrey, K.A., Mitchell, T.V., McKeown, M. J., Goldstein, J., Allison, T., McCarthy, G. (2003). Brain activity evoked by the perception of human walking: controlling for meaningful coherent motion. *Journal of Neuroscience*, 23(17), 6819–25.

Pereira, F., Mitchell, T. and Botvinick, M. (2009). Machine learning classifiers and fMRI: A tutorial overview. *Neuroimage*, 45 (1), 199-209.

Perry, A. & Bentin, S (2009). Mirror activity in the human brain while observing hand movements: a comparison between EEG desynchronization in the mu-range and previous fMRI results. *Brain Research, 1282*, 126-132.

Pineda, J. (2005). The functional significance of mu rhythms: Translating “seeing” and “hearing” into “doing”. *Brain Research Reviews, 50 (1),* 57-68.

Press, C., Gillmeister, H., & Heyes, C. (2007). Sensorimotor experience enhances automatic imitation of robotic action. *Proc Biol Sci,* 274(1625), 2509-2514.

Press, C., Cook, J., Blakemore, S. J. & Kilner, J. M. (2011). Dynamic modulation of human motor activity when observing actions. *Journal of Neuroscience, 31(8),* 2792-2800.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*(2), 131-141.

Rizzolatti, G., Fogassi, L. and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience, 2*, 661-670.

Saygin, A. P., & Cicekli, I. (2002). Pragmatics in human-computer conversations. *Journal of Pragmatics, 34(3)*, 227-258.

Saygin, A. P., Wilson, S. M., Hagler, D. J., Jr., Bates, E., & Sereno, M. I. (2004). Point-light biological motion perception activates human premotor cortex. *Journal of Neuroscience, 24(27)*, 6181-6188.

Saygin, A. P. (2007). Superior temporal and premotor areas are necessary for biological motion perception. *Brain, 130 (9),* 2452-2461.

Saygin, A. P., Chaminade, T., Urgen, B.A., & Ishiguro, H. (2011). Cognitive neuroscience and robotics: A mutually beneficial joining of forces. In L. Takayama (Ed.), Robotics: Systems andScience. Los Angeles, CA.

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J. & Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience, 7(4)*, 413-422.

Saygin, A. P., & Stadler, W. (2012). The role of appearance and motion in action prediction. *Psychological Research*, *76*, 388-394.

Shahin, A. J., Picton, T. W., & Miller, L. M. (2009). Brain oscillations during semantic evaluation of speech. *Brain and Cognition*, *70*(3), 259–66.

Silas, J., Levy, J. P., & Holmes, A. (2012). Sensitivity of ‘mu’rhythm modulation to the relevance of an observed movement but not to goal congruency. *International Journal of Psychophysiology, 85(2),* 168-173.

Sitnikova, T., Holcomb, P.J., Kiyonaga, K.A. & Kuperberg, G.R. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of Cognitive Neuroscience, 20*(11), 2037-2057.

Steckenfinger, S. A., & Ghazanfar, A. A. (2009). Monkey visual behavior falls into the uncanny valley. *Proceedings of the National Academy of Sciences of the United States of America, 106*(43), 18362-18366.

Steele, V. R., Bernat, E. M., van den Broek, P., Collins, P., Patrick, C. J. & Marsolek, C. J. (2013). Separable processes before, during, and after the N400 elicited by previously inferred and new information: Evidence from time–frequency decompositions. *Brain Research, 1492*, 92-107.

Stefan, K., Cohen, L. G., Duque, J., Mazzocchio, R., Celnik, P., Sawaki, L., Ungerleider, L., et al. (2005). Formation of a motor memory by action observation. *Journal of Neuroscience, 25*(41), 9339–46.

Tai, Y. F., Scherfler, C., Brooks, D. J., Sawamoto, N., & Castiello, U. (2004). The human premotor cortex is 'mirror' only for biological actions. *Current Biology, 14(2),* 117-120.

Thompson, J. C., Trafton, J. G. & McKnight, P. (2011). The perception of humanness from the movements of synthetic agents. *Perception, 40(6),* 695–704.

Tinwell, A., Grimshaw, M., Abdel-Nabi, D., Williams, A. (2011). Facial expression of emotion and perception of the uncanny valley in virtual characters. *Computers in Human Behavior, 27 (2)*, 741–749.

Tognoli, E., Lagarde, J., DeGuzman, G. C., & Kelso, J. S. (2007). The phi complex as a neuromarker of human social coordination. *Proceedings of the National Academy of Sciences, 104(19)*, 8190-8195.

Ugur, E., & Erol, S. (2011). Unsupervised learning of object affordances for planning in a mobile manipulation platform. *Robotics and Automation (ICRA) IEEE International Conference*.

Umiltà, M., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., & Rizzolatti, G. (2001). I know what you are doing: A neurophysiological study. *Neuron*, *31*(1), 155–65

Urgen, B.A., Plank, M., Ishiguro, H., Poizner, H., Saygin, A.P. (2012). Temporal dynamics of action perception: The role of biological appearance and motion kinematics. *34th Annual Conference of Cognitive Science Society*, Sapporo, Japan.

van Kemenade, B. M., Muggleton, N., Walsh, V. & Saygin, A.P. (2012). Effects of TMS over premotor and superior temporal cortices on biological motion perception. *Journal of Cognitive Neuroscience, 24(4),* 896-904.

Wermter, S., & Elshaw, M. (2003). Learning robot actions based on self-organising language memory. *Neural networks,* 16(5-6), 691–9.

Wu, Y.C. and Coulson, S. (2011). Are depictive gestures like pictures? Commonalities and differences in semantic processing. *Brain and Language, 119*(3), 184-195.

Zion-Golumbic, E., Kutas, M., & Bentin, S. (2009). Neural dynamics associated with semantic and episodic memory for faces : Evidence from multiple frequency bands. *Journal of Cognitive Neuroscience, 22(2)*, 263–277.

CHAPTER 3

Visual Action Population Codes in the Human Brain: Multivariate Pattern Analysis of fMRI Responses in Brain Areas involved in Visual Processing of Actions

Abstract

Over the last two decades neurophysiological and neuroimaging studies have identified a network of brain regions in occipito-temporal, parietal, and frontal cortex that are involved in visual processing of actions. What remain unclear are the neural computations and representational properties in each area. In this study, we investigated the representational content of human brain areas in the action observation network using fMRI and representational similarity analysis. Observers were shown video clips of 8 different actions performed by 3 different agents (actors) during fMRI scanning. We then derived two indices from the representational similarity matrices for each region of interest (ROI): Agent decoding index and action decoding index, which reflect the presence of significant agent and action information, respectively. We found significant agent decoding in early visual areas and category sensitive cortical regions including FFA and EBA, as well as in the core nodes of the action observation network. However, agent decoding index varied across ROIs and was strongest in the right posterior superior temporal sulcus (pSTS), and was significantly greater than the indices in ROIs in the parietal and frontal cortex. On the other hand, although we found significant action decoding in all visual areas as well as the core action observation network, the strength of action decoding was similar across ROIs. On the other hand, the representational structure of action types varied across ROIs as revealed by hierarchical clustering,

indicating that action-related information changes along the levels of the cortical hierarchy. These results suggest that during visual action processing, pSTS pools information from the early visual areas to compute the identity of the agent, and passes that information to regions in parietal and frontal cortex that code higher-level aspects of actions, consistent with computational models of visual action recognition.

3.1 INTRODUCTION

Action perception is one of the most important skills primates have, which is considered to be the building block of several higher level social skills such as communication, intention understanding and empathy (Blake and Shiffrar, 2007). Systems and cognitive neuroscience research in the last two decades have identified a dedicated system known as action observation network (AON) or mirror neuron system (MNS) in the primate brain that processes other individuals' actions (Rizzolatti and Craighero, 2004; Iacoboni and Dapretto, 2006; Vangeneugden et al., 2009; 2011; 2014; Nelissen et al., 2011). In its classical formulation, this system consists of areas in the occipito-temporal, parietal, and frontal cortices, namely posterior superior temporal sulcus (pSTS), inferior parietal lobule (IPL), and premotor cortex, respectively.

Among the nodes of the action observation network, pSTS is considered to be the area that gathers information from the visual cortex, which processes low-level visual stimulus features. Computational modeling suggests that pSTS integrates form and motion information from the ventral and dorsal pathways of the visual system, respectively, which constitute the two important cues in a perceived moving agent (Giese and Poggio, 2003). This information is then passed to the parietal and frontal nodes of the action observation network, which process high-level aspects of actions. Despite recent progress, we still lack a full understanding of the representations in each of these areas and how they differ from each other.

It is important to note that earlier studies exclusively focused on action observation network, more specifically on the brain areas that are considered to contain mirror neurons, and visual cortex has largely been neglected although it is well established that face and body stimuli, which make up an agent recruit areas in category-sensitive cortex such as EBA (Downing et al., 2001) and FFA (Kanwisher et al., 1997) as well as early visual areas. Furthermore, earlier studies generally focused on a specific region in the action observation network such as pSTS, parietal cortex or ventral premotor cortex mostly due to the limitations of the univariate analysis, which requires a lot of trials per condition, and thus experimenters had to use a small number of conditions, which would target a single region of interest.

In the present study, we addressed both of these limitations in an fMRI experiment with human subjects. First of all, we extend the investigation to all areas that are involved in visual processing of actions reported by experimental research and computational modeling (Giese and Poggio, 2003; Caspers et al., 2010). These areas include the classic action observation network as well as early visual cortex and category-sensitive cortex (Figure 3.1).

Second, we took advantage of representational similarity analysis (Kriegeskorte et al., 2008), a multivariate pattern analysis technique, which is ideal for condition-rich experiments. We scanned human subjects with fMRI by visually displaying a database of videos consisting of three agents and eight different actions, making up 24 different stimuli, and perform representational similarity analysis to make inferences about the representational properties of areas that process actions with respect to agent and action type as well as how the representations change along the different regions of the cortical hierarchy.

3.2 MATERIALS AND METHODS

3.2.1 Participants

27 subjects (12 females, 15 males) from the undergraduate and graduate student community at the University of California, San Diego participated in the study. Data of 4 subjects were not included in the data analysis due to large head movements (3 subjects) and technical problems in data acquisition (1 subject). The subjects had normal or corrected-to-normal vision and no history of neurological disorders. Informed consent was obtained in accordance with UCSD Human Research Protections Program. The subjects were paid $25 for 1.5 hours participation in the study.

3.2.2 Stimuli

Stimuli were video clips of actions performed by 3 agents: the humanoid robot Repliee Q2 in two different appearances (robotic and human-like appearances) and by the human ‘master’ after whom Repliee Q2 was modeled. We call these agents Robot, Android, and Human, respectively (also see Saygin et al. 2012 and Urgen et al. 2013 for additional details about the stimuli). The robot’s movement kinematics was mechanical differing from dynamics of biological motion. All the agents performed 8 different actions. The actions were comprised of a variety of transitive, intransitive, and communicative actions. Transitive actions included drinking from a cup, grasping an object, throwing a paper, and wiping a table; intransitive actions included nudging and turning to the right; and communicative actions included handwaving and talking (for introducing herself).

3.2.3 Procedure

Each participant was given exactly the same introduction to the study and the same exposure to the videos. Before starting fMRI scans, subjects were shown each video and were told whether each agent was a human or a robot (and thus were not uncertain about the identity of the agents during the experiment). We recorded fMRI BOLD response as subjects watched 2 sec video clips of the three agents performing eight different body actions (drinking from a cup, grasping an object, handwaving, talking, nudging, throwing a paper, turning to right, wiping a table). The experiment had a block design in which blocks consisted of video clips of one agent type (Human, Android, or Robot, see Figure 3.2). The experiment had 18 stimuli blocks (6 Human, 6 Android, 6 Robot) and they were presented in a pseudo-randomized order ensuring that all order combinations were presented (i.e. H-A-R, H-R-A, A-H-R, A-R-H, R-H-A, R-A-H). Presentation of three blocks of the agents was always followed by a rest block in which subjects fixated a cross for a time interval varying between 8.1 sec and 13.5 sec. Each block had 9 trials (8 different actions and repetition of a randomly chosen action once) with 0.1 sec inter-stimulus interval in between the trials. Each subject was presented a different order of blocks and of stimuli within each block. Subjects performed a 1-back task throughout the experiment by pressing a button whenever a movie was repeated in a block.

3.2.4 Localizers

In order to study the functional properties of several brain regions that are involved during visual action processing, we took a region of interest (ROI) approach and conducted localizer experiments. These areas include extra-striate body area (EBA), known to respond selectively to pictures of human bodies in contrast to other object categories (Downing, 2001); and fusiform face area (FFA), known to respond selectively to pictures of human faces in contrast to other object categories (Kanwisher et al., 1997).

EBA localizer: The localizer experiment to identify the functional area EBA consisted of blocks of pictures of human bodies and chairs (as an object category). There were 8 blocks for each stimulus type (body or chair). Presentation of two stimuli blocks was always followed by a rest block in which subjects fixated a cross at the center of the screen for a time interval varying between 14 sec and 18 sec. Each stimulus block consisted of 28 trials with 200 ms inter-stimulus interval in between the trials and 500 ms stimulus presentation. Subjects performed a 1-back task throughout the experiment by pressing a button whenever a stimulus was presented in each block.

FFA localizer: The localizer experiment to identify the functional area FFA consisted of blocks of pictures human faces and houses (as an object category). There were 8 blocks for each stimulus type (face or house). Presentation of two stimuli blocks was always followed by a rest block in which subjects fixated a cross at the center of the screen for a time interval varying between 14 sec and 18 sec. Each stimulus block consisted of 28 trials with 200 ms inter-stimulus interval in between the trials, and 500 ms stimulus presentation. Subjects performed a 1-back task throughout the experiment by pressing a button whenever a stimulus was presented in each block.

3.2.5 Image acquisition, preprocessing and first-level analysis

We scanned our subjects at the Center for fMRI at UC San Diego using the 3T scanner.

Main Experiment: First, the fMRI data of each subject were pre-processed with standard procedures including motion correction, slice-time correction, normalization, and smoothing using the SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). Then, two different first-level analyses were done using general linear model (GLM). In the first analysis, each agent type (Human, Android, Robot) as well as the rest blocks (fixation) were modeled as a separate condition and beta images were generated for these conditions. This analysis was done to identify the overall activity patterns and determine the ROIs of the action observation network. In the second analysis, each trial of all stimulus types was modeled separately and beta images corresponding to each trial were generated for each voxel. This analysis was done to prepare the single trials for Representational Similarity Analysis. Motion parameters generated in the preprocessing stage were used as regressors in both analyses.

Localizer scans: For each localizer experiment (EBA, FFA), the fMRI data of each subject were pre-processed with standard procedures including motion correction, slice-time correction, normalization, and smoothing. Then, using general linear model, beta images corresponding to each condition were generated for each voxel. The conditions corresponded to body and chair pictures in the EBA localizer, and face and house pictures in the FFA localizer. The condition differences were assessed at each voxel, and the voxels that showed reliable differences between conditions ($p < 0.001$ uncorrected) were used to identify the functional regions of interest (ROI) for each individual subject.

3.2.6 Identification of ROIs

We identified the ROIs of the action observation network by contrasting the overall activation patterns for all stimuli conditions compared to fixation ($p < 0.001$ uncorrected) using the first first-level analysis for each subject (described in *Main Experiment* section above). This contrast resulted in activation in early visual cortex, pSTS, parietal cortex and frontal cortex consistent with prior literature. Then, we chose the central voxel of the activation in the each of these areas and extracted a sphere ROI with a certain radius that covers the activation pattern. We did this for all subjects. As a result of this procedure, most of the subjects ended up having overlapping ROIs in each area of interest. However, several subjects had larger activation patterns, and using the central voxel approach resulted in these subjects having non-overlapping ROIs with the rest of the subjects. In order to cover the same areas and be consistent across subjects, we reduced the size of the ROIs of these subjects. This was done by moving the central voxel of the activation towards the +-15 mm of the average coordinates observed in the other subjects. This procedure is essentially similar to an AND operation for the activation in a certain area (e.g. parietal cortex) in all subjects to find the common areas of activation.

3.2.7 Representational similarity analysis (RSA)

In order to study the functional properties of the brain regions involved in action processing, and investigate the representational structure of the brain regions for actions, we used representational similarity analysis (Kriegeskorte et al., 2008). This technique allows researchers to quantify how similar the neural patterns that correspond to different conditions of an experiment within a certain brain region. In addition, application of clustering methods (e.g. multi-dimensional scaling) on the similarity measures allows one to study the representational structure of the particular brain region. Moreover, comparison of the similarity structures across brain regions allows one to study how neural representations change along the cortical hierarchy. With this method, it is also possible to compare the similarity of the neural representations with that of the stimuli, especially for visual cortex.

We calculated neural similarity matrices in each region of interest (ROI) for each subject by taking the correlation distance between all pairs for stimuli using the beta images derived in the first-level analysis in SPM8, which resulted in a 24 x 24 matrix. We computed the grand average similarity matrix by taking the average of all subjects for each ROI. To quantitatively compare the representations in each ROI, we derived two indices from the representational similarity matrices for each region of interest (ROI): Agent decoding index and action decoding index, which reflect the presence of significant agent and action information, respectively. Agent decoding index was calculated by subtracting the average off-diagonals from the average diagonals in the representational similarity matrix (Figure 3.5A). Action decoding index was calculated by subtracting the off-diagonals from the average diagonals in the representational similarity matrices (Figure 3.6A). We then ran 8 (ROI) X 2 (hemisphere) repeated-measures ANOVA on the Agent decoding index and Action decoding index to see whether the ROIs differ in terms of their agent and action discriminability. In order to closely examine the agent and action differences in each ROI, we constructed smaller representational similarity matrices only for agents by collapsing across actions (3 x 3 matrix since there are 3 agents) and only for actions by collapsing across agents (8 x 8 matrix since there are 8 actions) for each subject, and took the grand average of both matrix types separately across subjects. We then applied hierarchical clustering on the grand average dissimilarity matrices (1 – similarity matrix) to examine the similarity patterns for agents and actions. All the steps in representational similarity analysis were done using custom scripts in MATLAB (http://www.mathworks.com/products/matlab/).

3.3 RESULTS

3.3.1 Brain areas involved in visual processing of actions

Main Experiment: The visual stimuli in the main experiment resulted in activation in the early visual cortex (EVC), lateral occipital cortex (LOC), pSTS, parietal cortex, and frontal cortex, which were identified by running the GLM and contrast between all video stimuli and fixation in the main experiment ($p < 0.001$ uncorrected) (Figure 3.3 and Table 3.1).

Then, we extracted a sphere ROI that covered the activation pattern in each of these ROIs. We identified all ROIs in 17 subjects so the rest of the analysis included these subjects. The coordinates of the central voxels and the sizes of the spheres are provided in Table 3.2 for the ROIs in the visual cortex (EVC and LOC), and Table 3.3 for the ROIs that form the core nodes of the Action Observation Network.

Localizers: In EBA localizer, the contrast between images of body and chair stimuli resulted in activation patterns in EBA consistent with the coordinates of this area reported in the literature (Downing et al., 2001). Similarly, in FFA localizer, the contrast between images of faces and houses resulted in activation patterns in areas consistent with what is reported in the literature (Kanwisher et al., 1997). We then extracted a sphere ROI with a certain radius that covered the activation pattern in EBA and FFA (the coordinates of the central voxels and the sizes of the spheres are provided in Table 3.2 for each ROI).

3.3.2 Representational similarity analysis in the visual cortex and core nodes of the AON

The main structure of the representational similarity matrices (RSMs) is shown in Figure 3.4A, and the RSM for each region that was involved in visual processing of actions was shown in Figure 3.4B. RSA allows one not only to examine the representational structure within an ROI but it also allows one to examine how representations change along the cortical hierarchy across ROIs. Our RSA revealed that the ROIs in the visual cortex (EVC, LOC, EBA and FFA), and the core nodes of the AON (pSTS, parietal cortex, and premotor cortex) have qualitatively different representational patterns during perception of human and robot actions.

We then quantified some of the effects in the RSMs by computing Agent decoding index and Action decoding index, which are measures of significant agent information and action information, respectively.

3.3.3 Agent and Action Information in the ROIs

3.3.3.1 Agent decoding index

We computed the Agent decoding index by subtracting the across agent similarities from within agent similarities. This is done by subtracting the average off-diagonals (white entries) from the average diagonals (black entries) in the RSMs as shown in Figure 3.5A. An index that is significantly greater than zero reflects the presence of significant agent information.

All ROIs showed significant agent decoding (>0, one-sample t-test, $p<0.0001$), indicating that all ROIs could discriminate agents (Figure 3.5B). However, the strength of the discriminability changed across ROIs. It was strongest for pSTS and weakest for the ventral premotor cortex. 8 (ROI) x 2 (hemisphere) repeated measures ANOVA on the Agent decoding index showed a main effect of ROI ($F(7,112) = 2.16$, $p<0.05$). Post-hoc comparisons showed that ventral premotor cortex ($M = 0.044$, $SD = 0.005$) had significantly lower agent decoding index than all ROIs including LOC ($M = 0.064$, $SD = 0.006$) ($t(16) = -2.0$, $p<0.05$), EBA ($M = 0.053$, $SD = 0.006$) ($t(16) = -1.8$, $p<0.05$), FFA ($M = 0.060$, $SD = 0.008$) ($t(16) = -2.0$, $p<0.05$), pSTS ($M = 0.063$, $SD = 0.008$) ($t(16) = -2.5$, $p<0.05$), dorsal premotor cortex ($M = 0.058$, $SD = 0.006$) ($t(16) = -2.1$, $p<0.05$), except parietal cortex ($M = 0.052$, $SD = 0.005$) ($t(16) = -1.2$, $p = 0.1$). More importantly, one of the core nodes of the AON, pSTS had significantly higher agent discriminability than the other code node ventral premotor cortex (t(16) = -2.5, p<0.05). pSTS had marginally significant different from the other core node parietal cortex (t(16) = -1.7, p = 0.06). Other effects observed were LOC had significantly higher agent decoding index than the parietal cortex (t(16) = -2.4, p<0.05), and EBA had significantly lower agent decoding index than pSTS (t(16) = 1.9, p<0.05). There was not an effect of hemisphere (F(1,16) = 0.015, p = 0.9) or ROI x hemisphere interaction (F(7,112) = 0.76, p = 0.6).

3.3.3.2 Action decoding index

We computed the Action decoding index by subtracting the across action similarities from within action similarities. This is done by subtracting the average off-diagonals (white entries) from the average diagonals (black entries) in the RSMs as shown in Figure 3.6A. An index that is significantly greater than zero reflects the presence of significant action information.

All ROIs showed significant action decoding (>0, one-sample t-test, p<0.0001), indicating that all ROIs could discriminate actions (Figure 3.6B). However, the strength of the discriminability did not appear to change across ROIs. 8 (ROI) x 2 (hemisphere) repeated measures ANOVA on the Action decoding index did not show main effect of ROI (F(7,112) = 1.5, p = 0.2), of hemisphere (F(1,16) = 0.67, p = 0.4), or interaction effects (F(7,112) = 0.39, p = 0.9).

3.3.3.3 Representational geometry of agent and action information in the brain regions that are involved in visual processing of actions

After determining the discriminability of agents and actions in the ROIs of the AON, we further investigated the representational structure in each of these ROIs using agent-specific and action specific representational similarity matrices and hierarchical clustering. The basic aim of this approach was to identify which agents and actions show more similarities to each other. To this end, we constructed smaller similarity matrices for agents by collapsing across actions, and for actions by collapsing across agents. We then computed the dissimilarity matrix (1-similarity matrix) and applied hierarchical clustering on the dissimilarity matrices.

Agents: The representational similarity matrices and corresponding dendrograms (output of the hierarchical clustering) for agents were shown in Figure 3.7. As expected, we found that Human and Android had more similar neural representations in EVC compared to Robot since action clips with the Robot had a significantly different low-level visual differences than the other two agents. In LOC, EBA, and dorsal premotor cortex, we found that Human and Robot had more similar neural representations than that of Android. On the other hand, the neural representations of Human and Android were more similar to each other than Robot in FFA, pSTS, parietal cortex, and ventral premotor cortex.

Actions: The representational similarity matrices and corresponding dendrograms (output of the hierarchical clustering) for actions were shown in Figure 3.8. Our results show that although action discriminability did not change across ROIs as evidenced by the Action decoding index (see previous section), different ROIs have different representational structures and appear to discriminate actions based on different principles as can be clearly seen from the structure of dendrograms.

3.4 DISCUSSION

Although the last two decades of research has identified the brain regions that are involved in visual processing of action, the functional properties of these regions have remained largely unknown. In the present study, we investigated the functional properties of the brain regions that are involved in visual processing of actions using fMRI and representational similarity analysis. More specifically, we examined the representational properties of each region based on the agent and action type. This was novel because only a handful of human neuroimaging studies have focused on what aspects of an action each region was sensitive to (Jastorff et al., 2009; 2010; Saygin et al., 2012). Furthermore, unlike the previous studies on action observation, we extended our ROIs to visual cortex including early visual cortex and the category-sensitive cortex, which have been neglected in prior research. The motivation for doing that is to understand the neural processing during action perception starting from the visual cortex to the core nodes of the AON, similar to computational modeling approaches, which aim to provide a complete mechanism (Giese and Poggio, 2003). This was possible by means of employing the multivariate pattern analysis technique, representational similarity analysis (Kriegeskorte et al., 2008), which required a rich set of stimuli, instead of few stimuli required by univariate analysis techniques that could address questions restricted to a small number of brain regions. Although multivariate pattern analysis has been applied in many domains of cognitive neuroscience, its use in the action perception domain is relatively new (Filimon et al., 2014) and has been suggested to be a fruitful approach to improve our understanding of the action observation network (Oosterhof et al., 2013).

Our findings suggest that all brain regions that are involved in visual processing of actions including the regions in the visual cortex and the core nodes of the AON can discriminate between agents during visual perception of actions. However, the discriminability of agents differs across ROIs. As one goes higher in the visual cortex, the discriminability of agents increased, and it was highest in the pSTS, which is one of the core nodes of the AON that receives visual input from the visual cortex. Interestingly, the discriminability of agents decreased as one goes higher in hierarchy in the core nodes of the AON, i.e. from pSTS to ventral premotor cortex. This finding suggest that during visual processing of actions, pSTS might compute the identity of the agent based on the form and motion information it gets from the ventral and dorsal pathways of the visual system, and passes that information to the other core nodes of the AON that code for some higher level aspects of actions. This is consistent with a recent hypothesis that pSTS represents agent information during visual perception of actions (Orban, in press).

The representational geometry analysis that examined how agents were discriminated from each other, and the neural similarity of agents, showed that different regions had different representational patterns. In EVC, as expected agents that had similar low-level visual properties had similar neural representations. The neural representations of Human and Android were more similar to each other compared to Robot. In other nodes of the visual cortex including LOC and EBA, it appears that the congruence between the appearance and motion of the agent matters since the agents that had congruent appearance and motion (Human and Robot) had more similar neural representations than the agent that had incongruent appearance and motion (Android). The results for EBA were especially interesting since it is known to be sensitive to human bodies in contrast to other objects (Downing et al., 2001). Based on the functional properties of EBA, one could hypothesize that the agents that had more similar human-like bodies (Human and Android) would have more similar neural representations compared to the agent that had a less human-like body (Robot). The finding that Human and Robot had more similar representations suggests two alternative explanations. The first is that EBA might not be a region that is purely sensitive to visual form but rather might be involved in integration of form and motion information. The second alternative can be framed in terms of the predictive coding account of action perception (Kilner et al., 2007a; 2007b). In this account, every node in the cortical hierarchy sends a prediction error signal to the node in the lower node in the hierarchy. In the present study, it may be possible that due to the incongruence of appearance and motion information the Android had, there was a prediction error signal sent from a higher level in the hierarchy to EBA. On the other hand, in FFA, as expected the agents that had more similar human-like faces, Human and Android, had more similar neural representations compared to Robot, which has a less human-like face.

In the core nodes of the AON including pSTS, parietal cortex, and the ventral premotor cortex, the agents that had similar human-like appearance (Human and Android) had more similar neural representations compared to the agent that had less human-like appearance. The only node that had a different representational geometry was the dorsal premotor cortex in which the agents that had congruent appearance and motion (Human and Robot) had similar neural representations than the agent that had incongruent appearance and motion (Android). These results suggest that the core nodes of the AON except dorsal part of the premotor cortex have similar neural representations for agents that have similar human-like appearances compared to the agents that have less human-like appearances. Given the lack of evidence for similar neural representations based on motion suggest that in the core nodes of the AON, there is less sensitivity to the nature of the motion (biological vs. non-biological), and more sensitivity to human-like appearance or any other higher level processing based on human-like appearance.

Our findings about representation of different actions suggest that all brain regions that are involved in visual processing of actions including the regions in the visual cortex and the core nodes of the AON can discriminate between actions during visual perception of actions. The strength of discriminability of actions did not significantly differ across ROIs although it was strongest in the parietal cortex, which is hypothesized to have dedicated parts for different actions (Orban, in press). However, the representational geometry analysis that examined how actions were discriminated from each other revealed that different regions had different representational patterns, consistent with recent findings (Jastorff et al., 2010).

In conclusion, our study investigated the representational properties of the brain regions that are involved in visual processing of actions in the human brain with respect to the agent and action information in an observed action. Our findings suggest that different brain regions have different representational properties. More specifically, it appears that pSTS might compute the identity of the agent based on the form and motion information it gets from the ventral and dorsal pathways of the visual system, and passes that information to the other core nodes of the AON that code for some higher level aspects of actions, consistent with computational models of action perception (Giese and Poggio, 2003).

ACKNOWLEDMENTS

Chapter 3, in part, is currently being prepared for submission for publication of the material. Urgen, B.A. & Saygin, A.P. The dissertation author was the primary investigator and author of this paper. This study was supported by Qualcomm Institute, Kavli Institute for Brain and Mind, and DARPA. I would like thank my co-author Ayse P. Saygin. We also would like to thank Edward Nguyen for his help in fMRI data collection and programming the functional localizers.

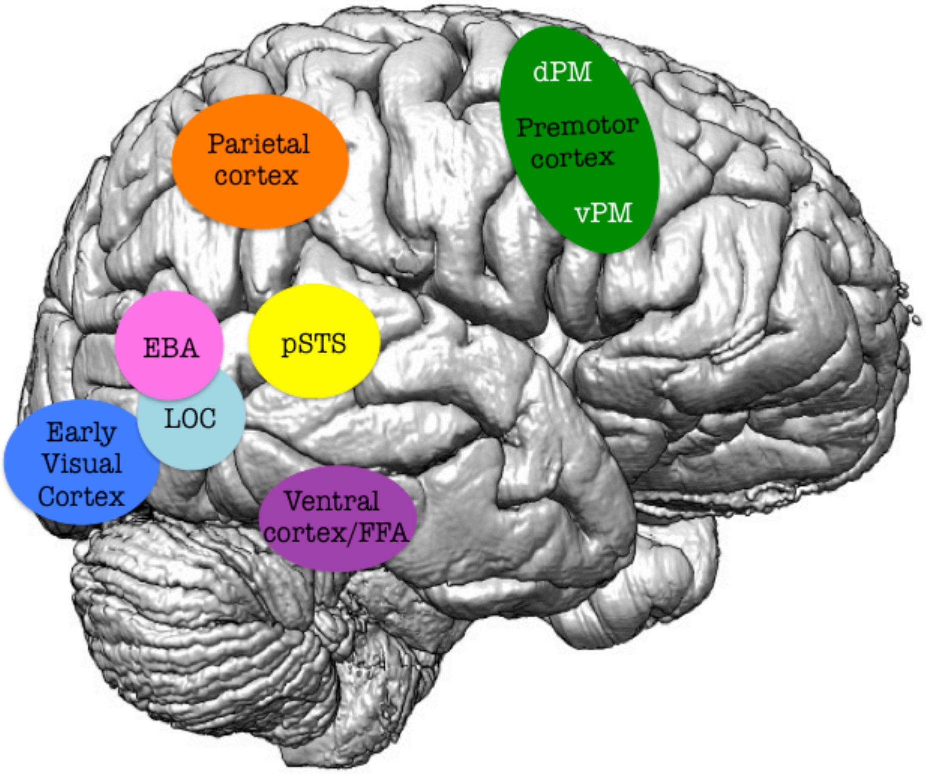

Figure 3.1: Human brain areas that are involved in visual perception of actions, including areas in the visual cortex and core nodes of the AON. Areas in the visual cortex include early visual cortex (EVC), lateral occipital cortex (LOC), extrastriate body area (EBA), and areas in the ventral cortex including fusiform gyrus (FFA). Areas in the classical formulation of the Action Observation Network include posterior superior temporal sulcus (pSTS), inferior parietal lobule (IPL), and dorsal (dPM) and ventral (vPM) premotor cortex.

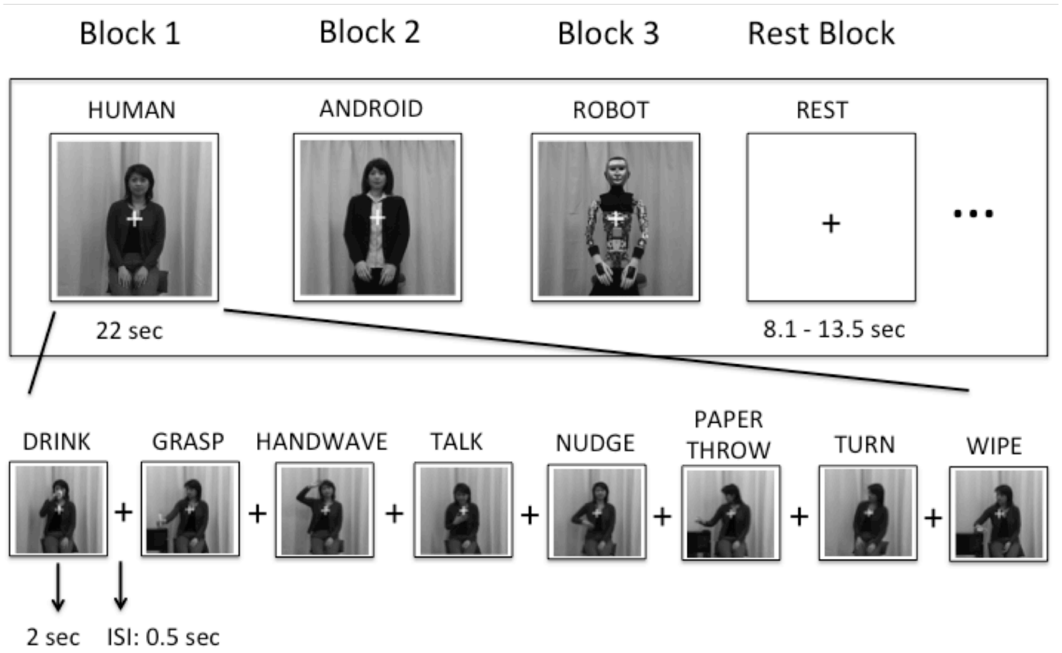

Figure 3.2: Experimental design of the fMRI study. It has a block design in which blocks consisted of video clips of one agent type (Human, Android, or Robot presented in a pseudo-randomized order with a rest block after each agent type is presented once. Each block consists of eight different actions and repetition of a randomly chosen action once, separated by ISI.

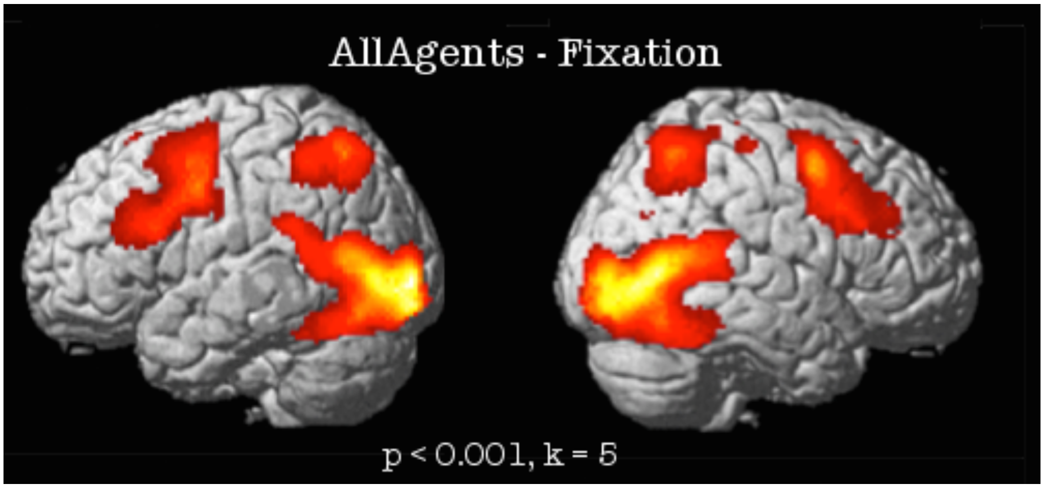

Figure 3.3: Whole brain GLM analysis with the contrast All Agents (Human, Android, Robot) – Fixation ($p < 0.001$, cluster threshold $k = 5$ voxels) across all subjects. The contrast revealed activation in early visual areas extending dorsally to lateral occipital cortex (LOC), and ventrally to inferior temporal cortex, pSTS, parietal cortex, and premotor cortex dorsally and ventrally in both hemispheres. See the coordinates in Table 3.1.

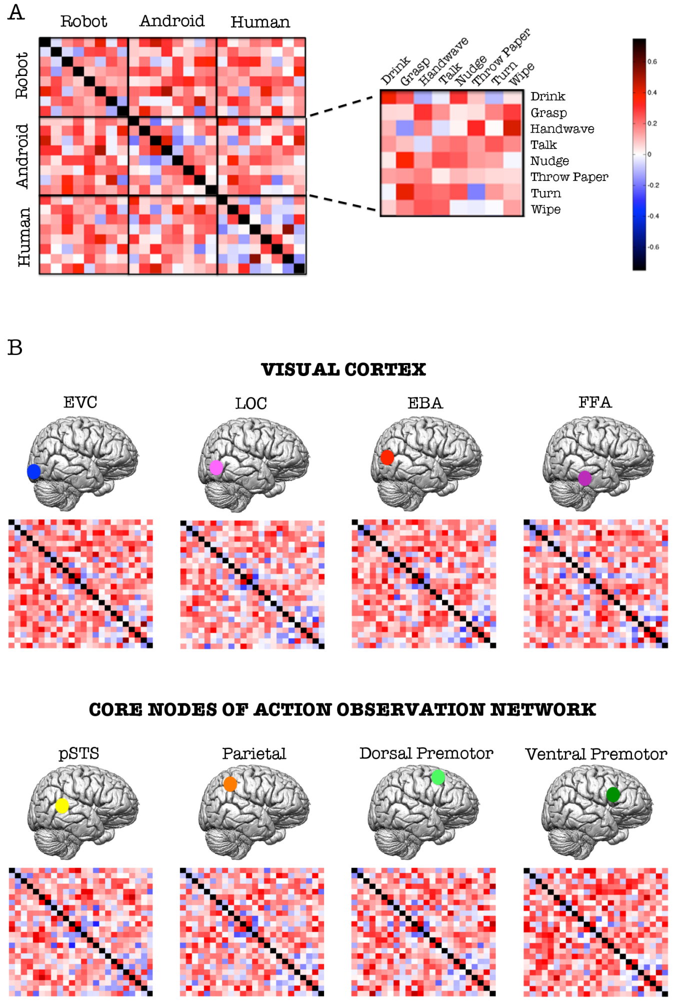

Figure 3.4: The structure of the representational similarity matrix, and the RSMs for each ROI of the AON. (A) The structure of the representational similarity matrices (RSM) for all ROIs. The RSMs are 24x24 matrices corresponding to pair-wise similarity of 3 agents (Robot, Android, Human), and 8 actions (Drink, Grasp, Handwave, Talk, Nudge, Throw Paper, Turn, Wipe). Each entry in the RSMs indicates the similarity of each pair of 24 stimuli. The similarity metric is Pearson's correlation. Note that the diagonals of RSMs indicate the similarity of each stimulus with itself and therefore are 1. (B) The RSMs of all ROIs that are involved in visual processing of actions. The ROIs including Early Visual Cortex (EVC), Lateral occipital cortex (LOC), pSTS, Parietal, Dorsal Premotor, and Ventral Premotor were all extracted for each individual subject from the All Agents-Fixation contrast in the whole brain GLM analysis (see Figure 3.3). EBA and FFA regions were extracted based on the EBA and FFA localizer scans. Top row shows the ROIs in the visual cortex and the bottom row shows the core nodes of the Action Observation Network.

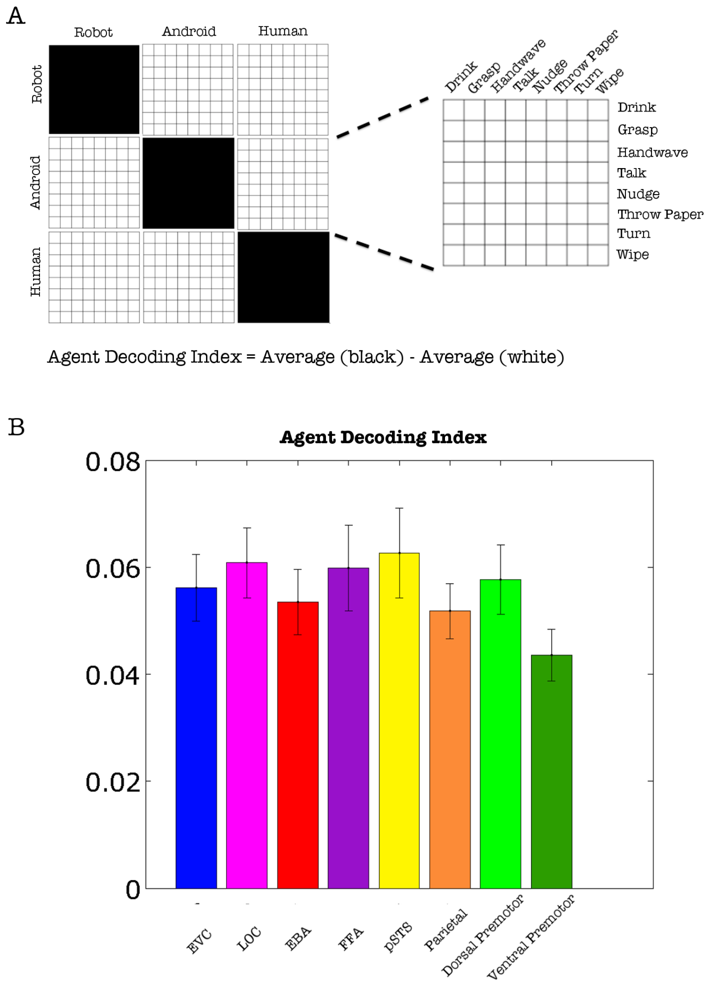

Figure 3.5: Agent Decoding Index across ROIs of the AON. (A) Computation of Agent Decoding Index from the representational similarity matrices. It is computed as the difference between the average of within agent similarities and across agent similarities. (B) Agent Decoding Index across ROIs. All ROIs had significant agent information (Agent Decoding Index > 0) evidenced by a one sample t-test ($p<0.0001$). In visual cortex (EVC, LOC, EBA, and FFA), Agent Decoding Index gradually increased from EVC to higher level visual areas including LOC and FFA suggesting that the discrimination of agents becomes stronger as one goes higher in the visual cortex. It was strongest at pSTS, which is one of the core nodes of the Action Observation Network that receives visual information from the visual cortex. In the core nodes of the Action Observation Network, Agent Decoding Index gradually decreased from pSTS to Ventral Premotor cortex suggesting that discrimination of agents becomes weaker as one goes higher in the hierarchy of the Action Observation Network. Agent Decoding Index at pSTS was significantly higher in Parietal cortex and Ventral Premotor cortex ($p<0.05$).

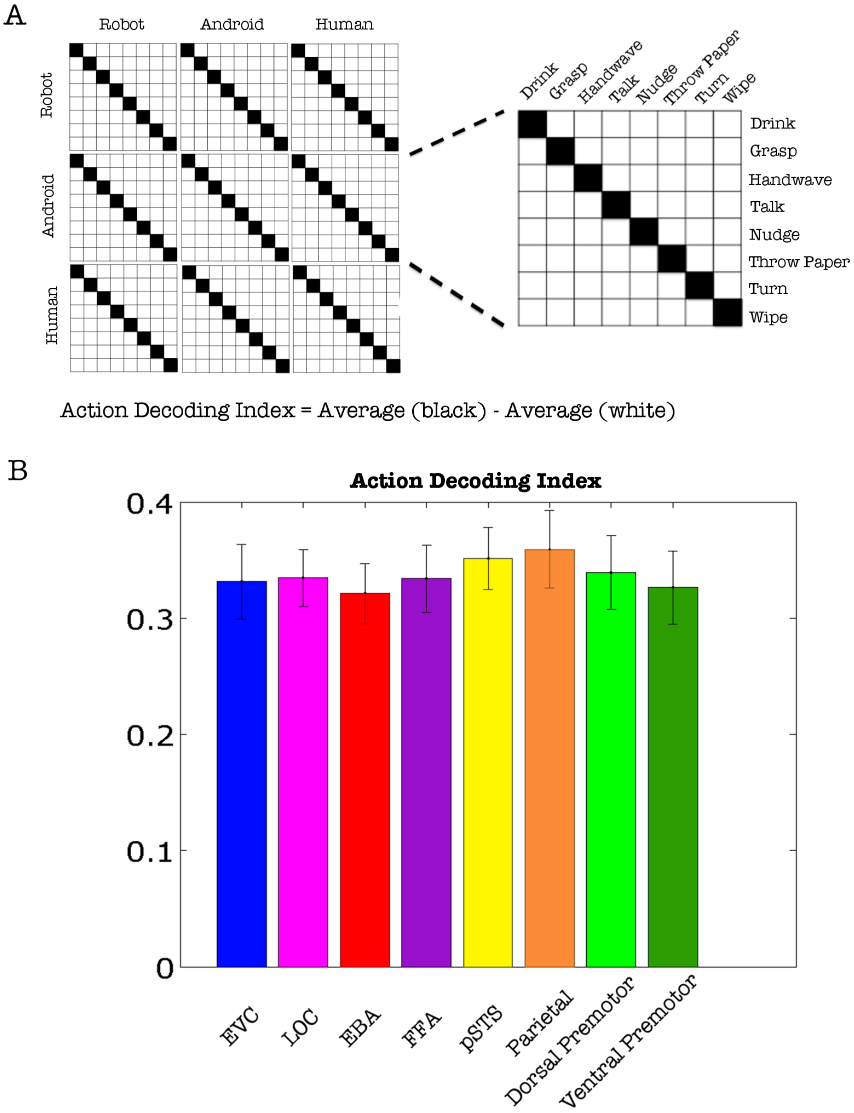

Figure 3.6: Action Decoding Index across ROIs of the AON. (A) Computation of Action Decoding Index from the representational similarity matrices. It is computed as the difference between the average of within action similarities and across action similarities. (B) Action Decoding Index across ROIs. All ROIs had significant action information (Action Decoding Index > 0) evidenced by a one sample t-test ($p < 0.0001$). In visual cortex (EVC, LOC, EBA, and FFA), Action Decoding Index did not differ between ROIs suggesting that discrimination of actions is similar in the visual cortex. In the core nodes of the Action Observation Network (pSTS, Parietal Cortex, and Premotor Cortex), it was strongest in the Parietal Cortex. It gradually increased from pSTS to Parietal Cortex and then decreased from Parietal Cortex to Premotor Cortex, suggesting that discrimination of actions becomes stronger as one goes from pSTS to Parietal Cortex but decreases as one goes from Parietal to Premotor Cortex. However, there was no main effect of ROI for the Action Decoding Index.

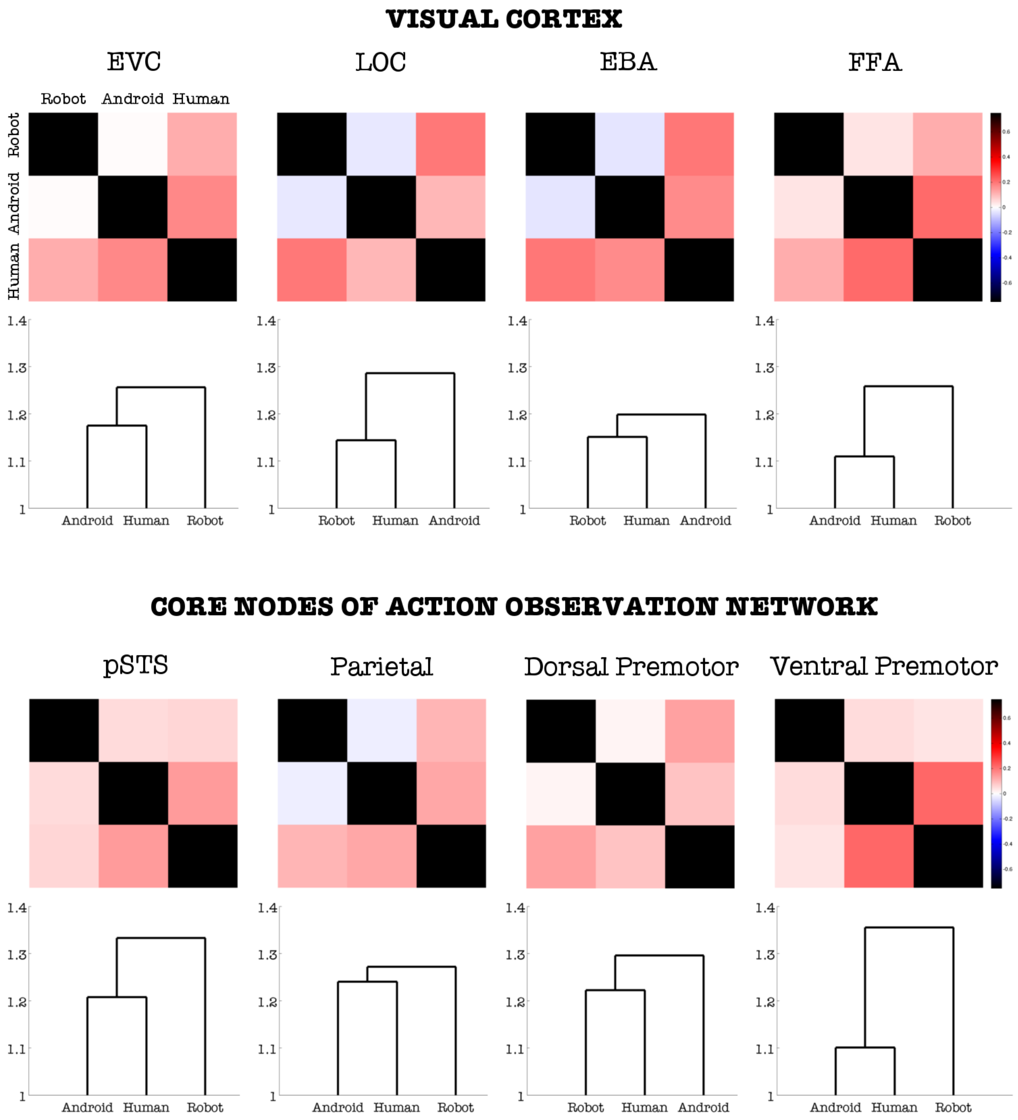

Figure 3.7: The representational similarity matrices (RSMs) and dendrograms only for Agents (Robot, Android, Human) after collapsing by actions in all ROIs. The top row shows the RSMs and dendrograms for the visual cortex, and the bottom row shows the RSMs and dendrograms for the core nodes of the Action Observation Network. There were two main representational geometries that emerged across ROIs: In EVC, FFA, pSTS, Parietal Cortex, and Ventral Premotor Cortex, Human and Android had more similar neural representations than Robot, whereas in LOC, EBA, and Dorsal Premotor Cortex, Human and Robot had more similar neural representations than Android.

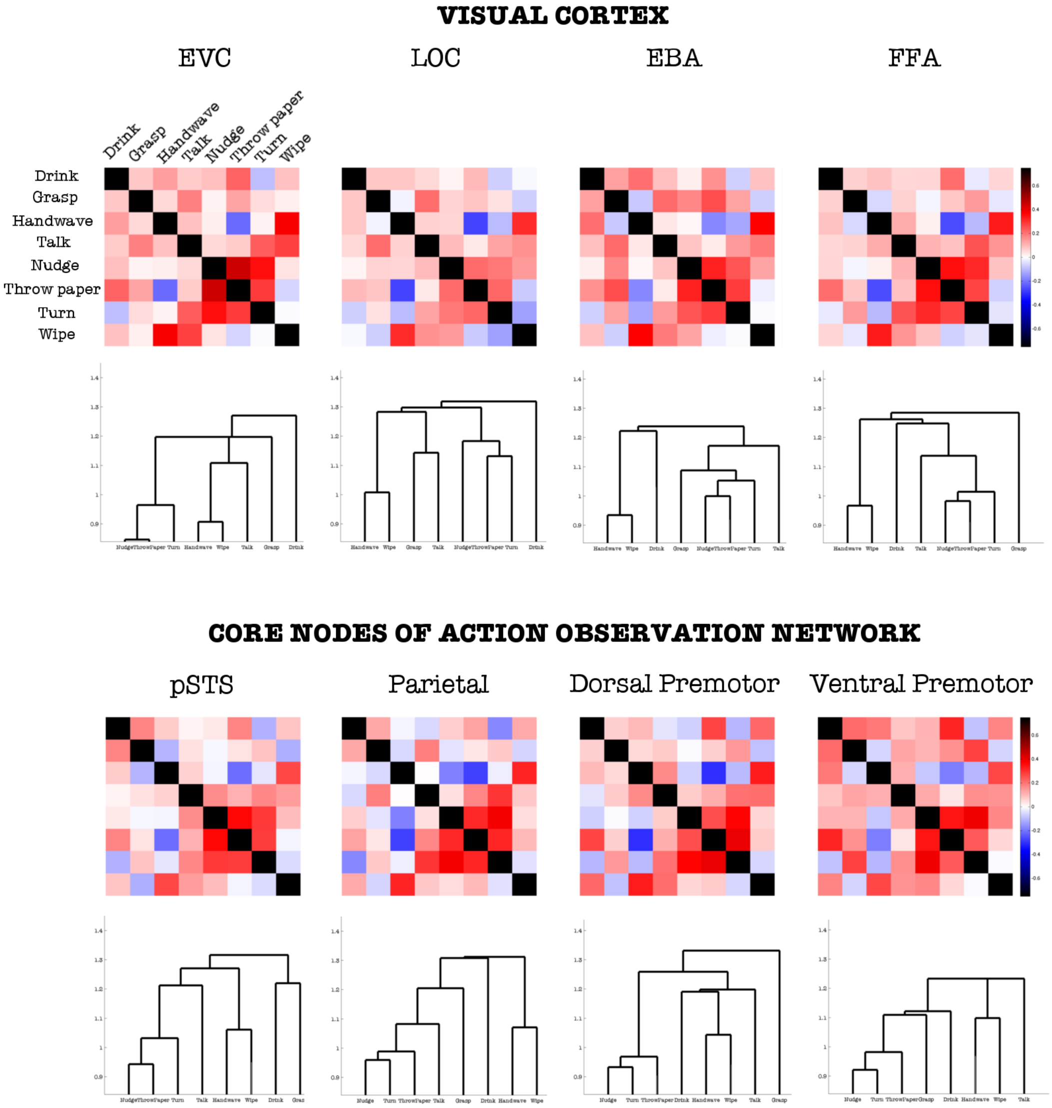

Figure 3.8: The representational similarity matrices (RSMs) and dendrograms only for Actions (Drink, Grasp, Handwave, Talk, Nudge, Throw Paper, Turn, Wipe) after collapsing by agents in all ROIs. The top row shows the RSMs and dendrograms for the visual cortex, and the bottom row shows the RSMs and dendrograms for the core nodes of the Action Observation Network. Although Action Decoding Index did not different between ROIs (see Figure 3.6), each ROI had a different representational geometry as can be seen from the RSMs and dendrograms, suggesting that although ROIs did not differ in their discriminability of actions based on Action Decoding Index, they discriminated actions based on different features.

Table 3.1: MNI Coordinates of the peak voxels of the brain regions involved in visual processing of actions based on the All Agents-Fixation contrast in the whole brain GLM analysis (Figure 3.3).

MNI coordinates				
x	**y**	**z**	**Anatomical Name**	**Brodmann Area**
-34	**-92**	**0**	Middle occipital gyrus (left)	BA 17
-26	-92	-10	Inferior occipital gyrus (left)	BA 18
-48	-80	-2	Middle occipital gyrus (left)	BA 19
48	**-74**	**-2**	Inferior temporal gyrus (right)	BA 19
40	-84	-8	Inferior occipital gyrus (right)	BA 19
22	-94	-6	Sub-gyral (right)	BA 18
42	**2**	**56**	Middle frontal gyrus (right)	BA6
50	34	34	Middle frontal gyrus (right)	BA9
46	10	30	Inferior frontal gyrus (right)	BA9
-34	**-58**	**50**	Superior parietal lobule (left)	BA39
38	**-56**	**52**	Inferior parietal lobule (right)	BA40
32	-68	28	Sub-gyral (right)	BA39
-44	**0**	**56**	Middle frontal gyrus (left)	BA6
-42	-2	38	Middle frontal gyrus (left))	
-60	6	32	Inferior frontal gyrus(left)	BA6
-6	**12**	**50**	Medial frontal gyrus (left)	BA6
28	**-6**	**-22**	Amygdala (right)	
-12	**26**	**60**	Superior frontal gyrus (left)	BA6
8	**-22**	**68**	Medial frontal gyrus (right)	BA6
38	**-26**	**58**	Precentral gyrus (right)	BA4

Table 3.2: Average MNI coordinates of the central voxels of the ROIs in the visual cortex together with the size of the spheres. The values in parenthesis indicate the standard deviation.

	VISUAL CORTEX					
	Left			**Right**		
	x y z	**Sphere size (radius in mm)**	**Anatomical name**	**x y z**	**Sphere size (radius in mm)**	**Anatomical name**
EVC	-23 -95 -1 (4 3 4)	8 (1)	Middle Occipital Gyrus	22 -95 2 (3 2 5)	8 (1.5)	Middle Occipital Gyrus
LOC	-47 -76 3 (3 4 5)	7 (0.9)	Middle Occipital Gyrus	50 -72 -1 (2 4 4)	8 (0.5)	Middle Occipital Gyrus
EBA	-48 -73 8 (3 6 7)	4 (0.5)	Middle Temporal Gyrus	53 -68 3 (3 6 5)	4 (0.7)	Middle Temporal Gyrus
FFA	-42 -51 -19 (3 7 5)	4 (0.4)	Fusform Gyrus	45 – 48 -22 (2 6 3)	4 (0.5)	Fusiform Gyrus

Table 3.3: Average MNI coordinates of the central voxels of the ROIs in the core Action Observation Network together with the size of the spheres. The values in parenthesis indicate the standard deviation.

	CORE NODES OF ACTION OBSERVATION NETWORK					
	Left			**Right**		
	x y z	**Sphere size (radius in mm)**	**Anatomical name**	**x y z**	**Sphere size (radius in mm)**	**Anatomical name**
pSTS	-50 -53 7 (5 9 4)	4 (0.9)	Superior Temporal Gyrus	53 -46 8 (5 7 4)	4 (0.8)	Superior Temporal Gyrus
Parietal	-31 -55 47 (4 5 7)	4 (1.5)	Inferior Parietal Lobule	35 -54 49 (6 5 6)	5 (0.8)	Superior Parietal Lobule
Dorsal Premotor	-43 -2 48 (4 3 6)	4 (0.8)	Precentral gyrus	45 3 48 (5 5 6)	4 (0.7)	Middle Frontal Gyrus
Ventral Premotor	-44 7 28 (4 7 5)	4 (0.5)	Inferior Frontal Gyrus	46 9 27 (5 9 4)	4 (0.7)	Inferior Frontal Gyrus

REFERENCES

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology*, *58*, 47–73.

Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *NeuroImage*, *50*(3), 1148–67.

Downing, P.E., Jiang, Y., Shuman, M. & Kanwisher, N. (1997). A cortical area selective for visual processing of human body. *Science, 293* (5539), 2470-3.

Filimon, F., Rieth, C.A, Sereno, M.I., & Cottrell, G.W. (2014). Observed, Executed, and Imagined Action Representations can be Decoded From Ventral and Dorsal Areas. *Cerebral Cortex, 25*(9):3144-58.

Harel, A., Kravitz, D. J., & Baker, C. I. (2014). Task context impacts visual object processing differentially across the cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(10), E962–71.

Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience*, *7*(12), 942–951.

Jastorff, J., & Orban, G. A. (2009). Human functional magnetic resonance imaging reveals separation and integration of shape and motion cues in biological motion processing. *Journal of Neuroscience, 29*(22), 7315–29.

Jastorff, J., Begliomini, C., Fabbri-Destro, M., Rizzolatti, G., & Orban, G. A. (2010). Coding observed motor acts: different organizational principles in the parietal and premotor cortex of humans. *Journal of Neurophysiology*, *104*(1), 128–40.

Jastorff, J., Popivanov, I. D., Vogels, R., Vanduffel, W., & Orban, G. A. (2012). Integration of shape and motion cues in biological motion processing in the monkey STS. *NeuroImage*, *60* (12), 911-21.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*(11), 4302-4311.

Kilner, J.M., Friston, K.J., Frith, C.D. (2007a) The mirror-neuron system: a Bayesian perspective. *Neuroreport, 18*:619-623.

Kilner, J.M., Friston, K.J., Frith, C.D. (2007b) Predictive coding: an account of the mirror neuron system. *Cognitive Processing*, *8*:159-166.

Kriegeskorte, N., Mur, M. and Bandettini, P. . (2008). Representational Similarity Analysis - Connecting the Branches of Systems Neuroscience. *Frontiers in Systems Neuroscience, 2*(4).

Nelissen, K., Borra, E., Gerbella, M., Rozzi, S., Luppino, G., Vanduffel, W., Rizzolatti, G. & Orban, G. A. (2011). Action observation circuits in the macaque monkey cortex. *Journal of Neuroscience*, *31*(10), 3743–56.

Oosterhof, N. N., Tipper, S. P., & Downing, P. E. (2013). Crossmodal and action-specific: neuroimaging the human mirror neuron system. *Trends in Cognitive Sciences*, 1–8.

Orban (in press). The mirror system in human and nonhuman primates: Comparative functional imaging studies suggest multiple systems.

Poggio, T., & Giese, M. A. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*, *4*(3), 179–92.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–92.

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. *Social Cognitive Affective Neuroscience, 7*(4), 413-422.

Urgen, B. A., Plank, M., Ishiguro, H., Poizner, H., & Saygin, A. P. (2013). EEG theta and Mu oscillations during perception of human and robot actions. *Frontiers in Neurorobotics*, *7*, 19.

Vangeneugden, J., Pollick, F., & Vogels, R. (2009). Functional differentiation of macaque visual temporal cortical neurons using a parametric action space. *Cerebral Cortex*, *19*(3), 593–611.

Vangeneugden, J., De Maziere, P. a., Van Hulle, M. M., Jaeggli, T., Van Gool, L., & Vogels, R. (2011). Distinct Mechanisms for Coding of Visual Actions in Macaque Temporal Cortex. *Journal of Neuroscience*, *31*(2), 385–401.

Vangeneugden, J., Peelen, M. V, Tadin, D., & Battelli, L. (2014). Distinct neural mechanisms for body form and body motion discriminations. *Journal of Neuroscience*, *34*(2), 574–85.

CHAPTER 4

Effective Connectivity in the Action Observation Network: Modulations by Agent Characteristics

Abstract

Visual perception of actions is supported by a network brain regions in the occipito-temporal, parietal, and premotor cortex in the human brain, known as Action Observation Network (AON). What remains largely unknown is the communication and direction of flow of information between these regions during perception of actions. According to predictive coding account of action perception, parietal regions of this network has reciprocal connections between the occipito-temporal regions such as pSTS and premotor cortex, and this network as a whole is not purely a feedforward system but rather has feedback connections through which prediction error signals are communicated between different regions. In the present study, we investigated the influence of two nodes of the AON, pSTS and premotor cortex over the third node, parietal cortex and how this influence was affected by a mismatch between the appearance and motion of an observed agent while performing an action, which hypothetically generates a prediction error, using fMRI and Dynamical Causal Modeling (DCM). Our DCM results suggest that the influence on parietal node is likely through a feedback connection from premotor cortex during perception of actions that violate people's predictions.

4.1 INTRODUCTION

Over the last two decades, neurophysiological and neuroimaging studies in primates have identified a network of brain regions in occipito-temporal, parietal and frontal cortex that are associated with perception of actions, known as the Action Observation Network (AON, Rizzolatti & Craighero, 2004; Iacoboni & Dapretto, 2006). Although the brain regions that are involved in the processing of actions have been identified, the communication between these regions has remained largely unknown. Our knowledge of the anatomical connectivity patterns in the action observation network comes primarily from non-human primates. In the macaque monkey, area F5 of the premotor cortex and area PF of the inferior parietal lobule (two key nodes of the AON that contain mirror neurons) have reciprocal connections (Luppino, et al., 1999). PF also has reciprocal connections with a portion of the posterior superior temporal sulcus (pSTS) that is sensitive to biological movements (Seltzer & Pandya, 1994). Analogous connectivity patterns have been proposed in the human brain (Rushworth et al., 2006). These anatomical findings have allowed researchers to develop models for functional connectivity within the AON.

One of the most prominent models developed by Kilner et al. (2007a; 2007b) proposes that the AON is a predictive system, following the principles of predictive coding (Friston, 2010). In this framework, in contrast to the classic formulation of the AON, which sees action perception strictly as a feedforward process, Kilner et al. (2007a; 2007b) propose that visual action information is processed throughout the AON by means of the reciprocal connections (i.e., both feedforward and feedback) between the pSTS and parietal cortex, and parietal cortex and the premotor cortex (Figure 4.1). In this framework, incoming information is compared with predictions at each level of the AON.

There is in fact recent experimental evidence that provides support for predictive coding account of action perception. Kilner et al. (2004), using event-related brain potentials, found that during action observation, the human brain generated a motor-preparation-like negative potential when the action was in a predictable context; no such potential was found when observation occurred within an unpredictable context. In a monkey neurophysiology study, Maranesi et al. (2014) provide direct evidence for predictive activity of mirror neurons and therefore it is considered to be a foundational step in supporting the predictive coding account of action understanding (Urgen and Miller, 2015). In another study, using an fMRI-adaptation paradigm, Saygin et al. (2012) found that the parietal node of the AON showed more adaptation to mis-predictable actions (via an agent who showed a mismatch between appearance and motion) than to corrently-predictable ones (via agent who shows a match between appearance and motion). The authors interpreted the differential adaptation in the parietal cortex for the mis-predictable actions as reflecting prediction error signals generated due to a mismatch between the appearance and movement of the observed actor.

Due to the activation-based analysis in Saygin et al. (2012), it could not be determined whether the influence on parietal cortex came as feedforward (bottom-up) modulation from earlier visual areas via pSTS, or as feedback (top-down) modulation from premotor cortex in AON in the mismatch condition. The current study aims to reveal where that influence to parietal cortex might be generated from. Is it a top-down signal from premotor cortex or a bottom-up signal from pSTS? To address this question, we studied the effective connectivity patterns in the action observation network of the human brain and their modulation by the agent characteristics using functional magnetic resonance imaging (fMRI) and dynamical causal modeling (DCM) (Friston et al., 2003). Specifically, we investigated the influence of two nodes of the AON, pSTS and premotor cortex over the third node, parietal cortex and how this influence was affected by congruence between the appearance and motion of the observed agent.

4.2 MATERIALS AND METHODS

4.2.1 Participants

27 subjects (12 females, 15 males) from the undergraduate and graduate student community at the University of California, San Diego participated in the study. Data of 4 subjects were not included in the data analysis due to large head movements (3 subjects) and technical problems in data acquisition (1 subject). The subjects had normal or corrected-to-normal vision and no history of neurological disorders. Informed consent was obtained in accordance with UCSD Human Research Protections Program. The subjects were paid $25 for 1.5 hours participation in the study. All ROIs of interest for DCM analysis were identified in 18 subjects so those subjects were included in the DCM analysis.

4.2.2 Stimuli

Stimuli were video clips of actions performed by 3 agents: the humanoid robot Repliee Q2 in two different appearances (robotic and human-like appearances) and by the human 'master' after whom Repliee Q2 was modeled. We call these agents Robot, Android, and Human, respectively (also see Saygin et al. 2012 and Urgen et al. 2013 for additional details about the stimuli). The robot's movement kinematics was mechanical differing from dynamics of biological motion. All the agents performed 8 different actions. The actions were comprised of a variety of transitive, intransitive, and communicative actions. Transitive actions included drinking from a cup, grasping an object, throwing a paper, and wiping a table; intransitive actions included nudging and turning to the right; and communicative actions included handwaving and talking (for introducing herself).

4.2.3 Procedure

Each participant was given exactly the same introduction to the study and the same exposure to the videos. Before starting fMRI scans, subjects were shown each video and were told whether each agent was a human or a robot (and thus were not uncertain about the identity of the agents during the experiment). We recorded fMRI BOLD response as subjects watched 2 sec video clips of the three agents performing eight different body actions (drinking from a cup, grasping an object, handwaving, talking, nudging, throwing a paper, turning to right, wiping a table). The experiment had a block design in which blocks consisted of video clips of one agent type (Human, Android, or Robot, see Figure 3.2). The experiment had 18 stimuli blocks (6 Human, 6 Android, 6 Robot) and they were presented in a pseudo-randomized order ensuring that all order combinations were presented (i.e. H-A-R, H-R-A, A-H-R, A-R-H, R-H-A, R-A-H). Presentation of three blocks of the agents was always followed by a rest block in which subjects fixated a cross for a time interval varying between 8.1 sec and 13.5 sec. Each block had 9 trials (8 different actions and repetition of a randomly chosen action once) with 0.1 sec inter-stimulus interval in between the trials. Each subject was presented a different order of blocks and of stimuli within each block. Subjects performed a 1-back task throughout the experiment by pressing a button whenever a movie was repeated in a block.

4.2.4 Image acquisition, preprocessing and first-level analysis

We scanned our subjects at the Center for fMRI at UC San Diego using the 3T scanner. First, the fMRI data of each subject were pre-processed with standard procedures including motion correction, slice-time correction, normalization, and smoothing using the SPM8 software. Then, two different first-level analyses were done using general linear model (GLM). In the first analysis, each agent type (Human, Android, Robot) as well as the rest blocks (fixation) were modeled as a separate condition and beta images were generated for these conditions. This analysis was done to identify the overall activity patterns and determine the ROIs of the action observation network. In the second analysis, we defined two conditions: The first condition was defined as *actions*, and consisted of all three agents (Human, Android, Robot). The second condition was defined as the *mismatch* condition, and consisted of the agent Android, as it featured a mismatch between appearance and motion of the agent. This second analysis was done to investigate the modulations in the action observation network. Motion parameters generated in the preprocessing stage were used as regressors in both analyses.

4.2.5 Identification of ROIs

We identified the ROIs of the action observation network by contrasting the overall activation patterns for all stimuli conditions compared to fixation ($p < 0.001$

uncorrected) using the first first-level analysis for each subject (described in Section 4.2.5 above). This contrast resulted in activation in visual cortex, pSTS, parietal cortex and premotor cortex consistent with prior literature. Then, we chose the central voxel of the activation in the each of these areas except visual cortex (since it was not included in the DCM analysis) and extracted a sphere ROI that covers the activation pattern. We did this for all 18 subjects for whom we identified all ROIs of interest. As a result of this procedure, most of the subjects ended up having overlapping ROIs in each area of interest. The ROI time series data was then extracted using eigenvariate (threshold of $p<0.05$) with a 4 or 5 mm radius of the sphere depending on the range of the activation pattern.

4.2.6 Specification of network models

Dynamical causal modeling (DCM) is an effective connectivity technique to estimate the directed functional connectivity, also known as effective connectivity patterns between different regions of interest with fMRI (Friston et al., 2003; Penny et al., 2004). The basic idea behind DCM is that it treats the brain as a deterministic non-linear system which is open to perturbations by external stimuli and which produces outputs. In this framework, the modeling approach infers hidden (unobserved) states from measured BOLD data via a generative model. This is done by formalizing the interaction between the hidden states in neurophysiologically plausible differential equations and transforming them into BOLD signal via a hemodynamic forward model.

DCM consists of two stages: Model specification and estimation, and model selection. In the model specification and estimation stage, several model architectures are specified based on the known anatomy between brain regions of interest and researcher's hypothesis about how the connections might be influenced by experimental manipulations. It then estimates three parameters: 1) Intrinsic connections between brain regions, 2) How the intrinsic connections are modulated by experimental manipulations, 3) The extrinsic input strength into the system. In the model selection stage, Bayesian Model Selection (BMS) procedure is used to determine the most likely model that generated the observed data. In this procedure, each model architecture in the model space tested by the researcher is given a probability for explaining the observed data. The model that has the highest probability is then considered to be the "winning" or the most optimal model, and the parameter estimates determined in the first stage for that model are reported.

To test our hypothesis, we constructed three models that consisted of the main three ROIs of the AON, namely pSTS, the parietal node, and the premotor node (specified in Section 4.2.5). To constrain the model space, in each of these models, the intrinsic connections between the ROIs were informed by the known anatomical connections between the regions. As such, pSTS and the parietal node, and the parietal node and the premotor node had reciprocal connections between each other. In all models, pSTS was considered to be the node where the visual input entered the system, and all intrinsic connections were modulated by the observation of actions (Figure 4.2A). In other words, the observation of actions was assumed to evoke activity in pSTS first (input to the system), and then subsequently propagated to parietal and premotor cortex based on the known anatomical connections. After the first feedforward flow of information from a lower area to a higher area, feedback from a higher area to a lower area occurred in the models.

The three models differ with regard to which connections are modulated by the mismatch condition (Figure 4.2B). The first model posits that influence on parietal cortex activity is through connections from pSTS to parietal cortex, i.e., a bottom-up modulation (Model 1). The second model posits that the influence on parietal cortex is through feedback from premotor cortex, i.e., a top-down modulation (Model 2). A third possibility is that the influence would be expressed through both pSTS and premotor cortex connections (Model 3).

To identify the winning model, i.e. the model that explains that data best, Bayesian Model Selection (BMS) was used. This method determines a probability for each model, known as the exceedance probability, which is the probability that a model is more likely than any other model tested in the model space.

4.3 RESULTS

4.3.1 Brain Regions that are involved in Visual Processing of Actions

In order to identify the brain regions that were involved in visual processing of actions, we ran the contrast between the observation of all agents (Human, Android, Robot) and the fixation condition. This contrast revealed the activation in early visual cortex extending dorsally to lateral occipital cortex (LOC) and ventrally to the inferior temporal cortex, as well as the core areas for action observation network, namely pSTS, parietal regions in the anterior part of the intra-parietal sulcus (AIP) and inferior and superior parts of the parietal lobe, and dorsal and ventral parts of the premotor cortex, all bilaterally ($p<0.001$) (Figure 4.3, also see Table 3.1 in Chapter 3 for the coordinates).

In order to deal with the expansion of model space with increasing number of ROIs and constrain the model space used in the effective connectivity analysis, we extracted ROIs from pSTS, parietal, and premotor cortex in each subject, and excluded the areas in early visual areas. The coordinates of the central voxels of the ROIs averaged over subjects are displayed in Table 4.1.

4.3.2 Effective Connectivity with DCM and Model Selection with BMS

The DCMs were created using the ROIs listed in Table 4.1, namely pSTS, the parietal node (AIP), and the premotor node (ventral premotor cortex). The intrinsic connections were reciprocal between pSTS and the parietal node, and the parietal node and the ventral premotor cortex. The input into the system was considered to come from pSTS. Action observation was considered to modulate all intrinsic connections (defined by the *action* condition, see Section 4.2.5), and the *mismatch* condition was considered to modulate either the pSTS-parietal connection (Model 1), premotor-parietal connection (Model 2), or both of these connections (Model 3).

BMS analysis on the three DCMs has shown that Model 2 was the winning (optimal) model in both hemispheres, whose probability was 0.43 in the left and 0.50 in the right (Figure 4.4). The next best model was Model 1 whose probability was 0.32 on the left, and 0.36 on the right. The least likely model, Model 3 had a probability of 0.25 on the left, and 0.14 on the right.

The intrinsic connection strengths between the ROIs in the winning model, Model 2, are listed in Table 4.2, and results for right hemisphere are shown in Figure 4.5. Importantly, all connection strengths were found to be greater than 0 by a one-sample t-test ($p<0.0001$, one sample t-test). Also, the reciprocal connections between the pSTS and the parietal node were estimated to be stronger than the other three connections both in the left and the right hemisphere.

The modulatory effect of observation of actions in Model 2 is shown in Table 4.3 and results of the right hemisphere are shown in Figure 4.6 together with the modulatory effect of mismatch condition and the input strength. Modulatory effects of actions on all four connections were found to be greater than 0 by a one-sample t-test ($p<0.0001$). The connection between pSTS and the parietal node appeared to be modulated most strongly. The modulatory effect of the mismatch condition on the premotor-parietal connection was 0.0003 on both hemispheres. The input strength was -0.0058 on the left hemisphere, and -0.0057 on the right hemisphere.

4.4 DISCUSSION

In the current study, we aimed to estimate the effective connectivity patterns between the core nodes of AON and how these connections were modulated by the characteristics of the agent observed. Our study was primarily motivated by the findings of Saygin et al. (2012), who reported that the parietal node of the AON (intra-parietal sulcus) showed differential activity during observation of actions which were performed by an agent who possessed a mismatch between appearance and motion (a biological appearance but mechanical motion) compared to other agents who possessed a match between appearance and motion (biological appearance and motion or mechanical appearance and motion). Within the predictive coding account of action perception (Kilner et al. 2007a; 2007b), one question that has been of interest to us was whether that differential activity was a top-down effect from premotor cortex or a bottom-up effect from pSTS. The current study addressed this question using fMRI and DCM.

We constructed three models to test our hypothesis in both hemispheres for each subject. First of all, informed by well-known anatomy, in all these models, we constructed reciprocal intrinsic connections between pSTS and parietal node, and the parietal node and the premotor node. The input into the system was considered to enter from pSTS, which was a reasonable assumption given the anatomical connectivity of pSTS between dorsal and ventral pathways of the visual system. In addition, in all these models, we assume that observation of actions modulate all intrinsic connections. We when constructed the three models that correspond to our hypotheses: A model in which the connection from pSTS to parietal cortex was modulated, a second model in which the connection between the premotor cortex and the parietal cortex was modulated, and a third model in which both connections were modulated by the mismatch condition (the agent that exhibited an incongruence between appearance and motion). Our results show that the most likely model that best explains the data is a model in which the connection between the premotor cortex to the parietal cortex was modulated, which indicates a top-down influence.

Examination of the parameter estimates of the optimal model shows that all of the intrinsic connections were different from 0, confirming the well-known anatomy between these regions. The strongest intrinsic connectivity was between pSTS and the parietal node. All of these intrinsic connections were modulated significantly by the observation of actions, although the strongest modulation was in the pSTS-parietal node connection. Although the model in which premotor to parietal connection was modulated by the mismatch condition was the optimal model based on the BMS, it must be noted that the modulatory strength of mismatch was weak in both hemispheres.

These results provide support for the predictive coding account of action perception (Kilner et al., 2007a; 2007b). It appears that the human brain processes actions within the AON via both feedforward and feedback connections. During the observation of actions, which were performed by an agent who exhibit a mismatch between appearance and motion, the premotor node of the AON seems to send a feedback signal to the parietal node in the lower part of the hierarchy, which might indicate a prediction error signal.

However, several caveats must be noted. First of all, BMS is a Bayesian approach, which assigns a probability to each model tested in the model space of the researcher, and having a winning model does not preclude the possibility that the other models are not likely. In our study, that the probability of the winning model is around 0.5, which is not very high relative to the other competing models tested. Second, although significant, the modulatory effect of the mismatch condition was found to be weak. These suggest that it is worth expanding the model space by adding more nodes (e.g. visual areas or even prefrontal areas that are known to influence premotor activity) when constructing the models of the AON.

ACKNOWLEGMENTS

Chapter 4, in part, is currently being prepared for submission for publication of the material. Urgen, B.A. & Saygin, A.P. The dissertation author was the primary investigator and author of this paper. This study was supported by Qualcomm Institute, Kavli Institute for Brain and Mind, and DARPA. I would like to thank my co-author Ayse P. Saygin. The authors also would like to thank Edward Nguyen for his help in fMRI data collection.

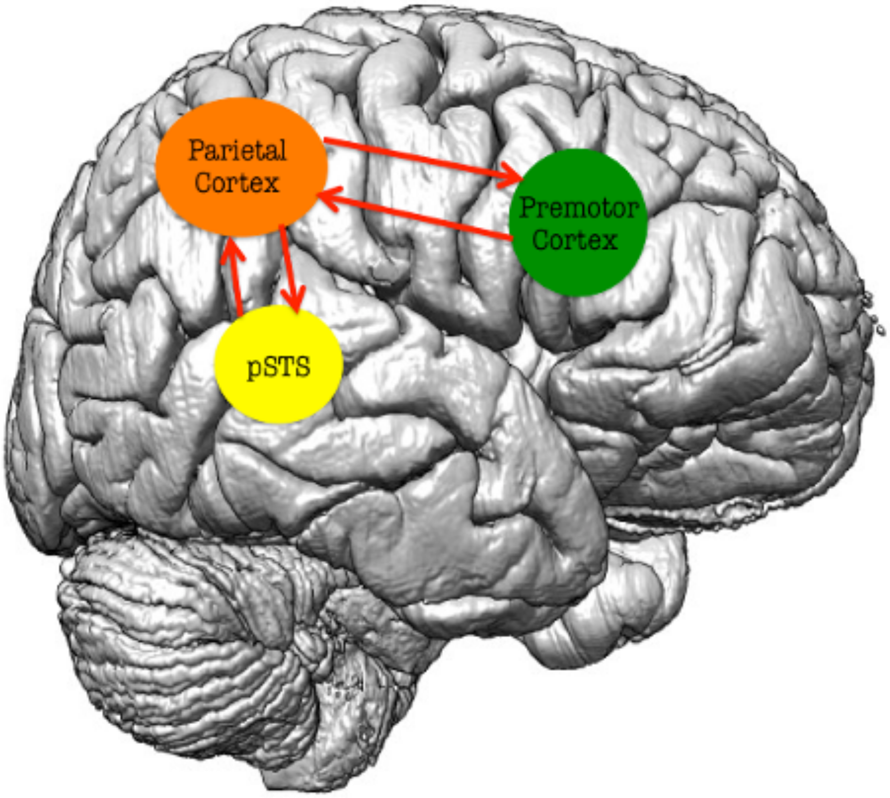

Figure 4.1: Anatomical connectivity between the core nodes of the AON.

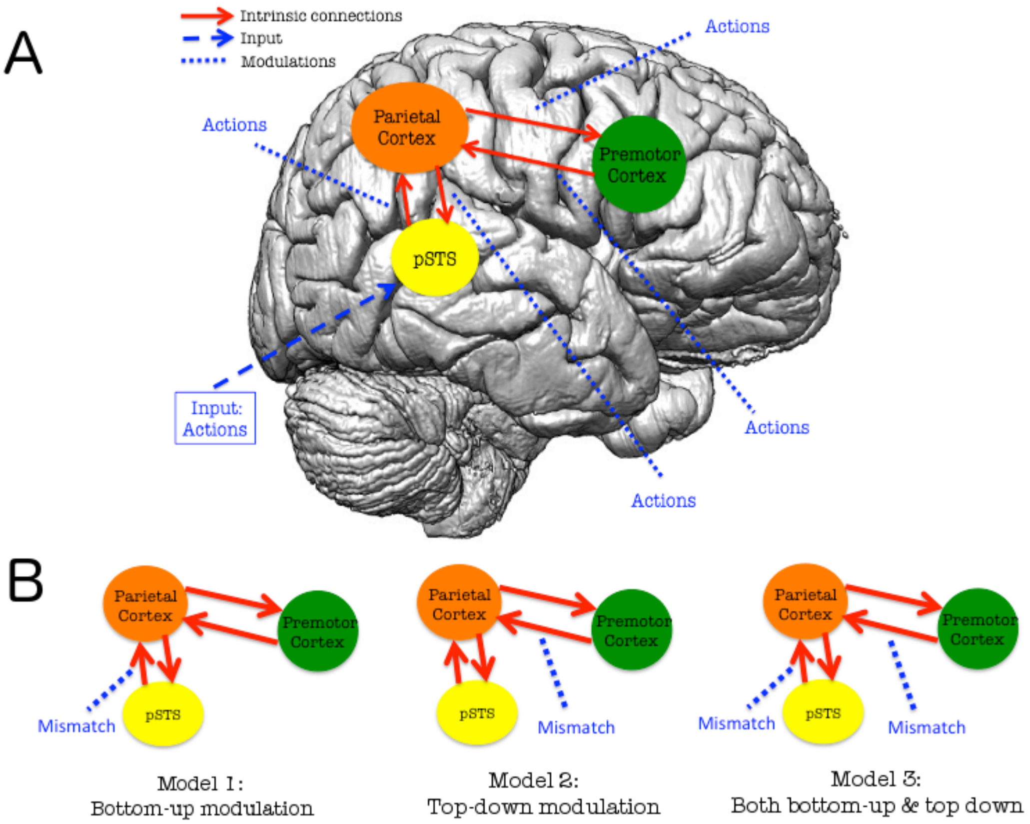

Figure 4.2: DCM models tested in the model space. (A) The DCM model that forms the basis for all tested models in the model space in (B). There are reciprocal intrinsic connections between pSTS and parietal node, and the parietal node and the premotor node informed by anatomy (red arrows). The input to the system is assumed to enter to the AON through pSTS since pSTS gets information from the visual cortex (blue dashed arrow). All the intrinsic connections are modulated by observation of actions (dashed blue lines). (B) The model space that consists of three models that correspond to our hypothesis about how the mismatch condition might modulate the intrinsic connections. Model 1 tests a bottom-up modulation from pSTS to parietal node, Model 2 tests a top-down modulation from premotor node to parietal node, and Models 3 tests both a bottom-up and a top-down modulation.

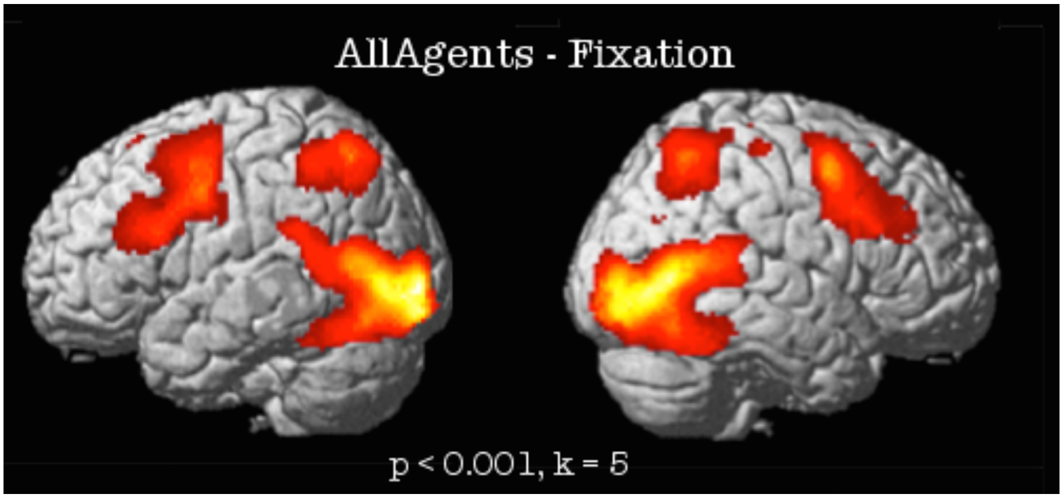

Figure 4.3: Whole brain GLM analysis with the contrast All Agents (Human, Android, Robot) – Fixation ($p < 0.001$, cluster threshold $k = 5$ voxels) across all subjects. The contrast revealed activation in early visual areas extending dorsally to lateral occipital cortex (LOC), and ventrally to inferior temporal cortex, pSTS, parietal cortex, and premotor cortex dorsally and ventrally in both hemispheres. See the coordinates in Table 3.1 in Chapter 3.

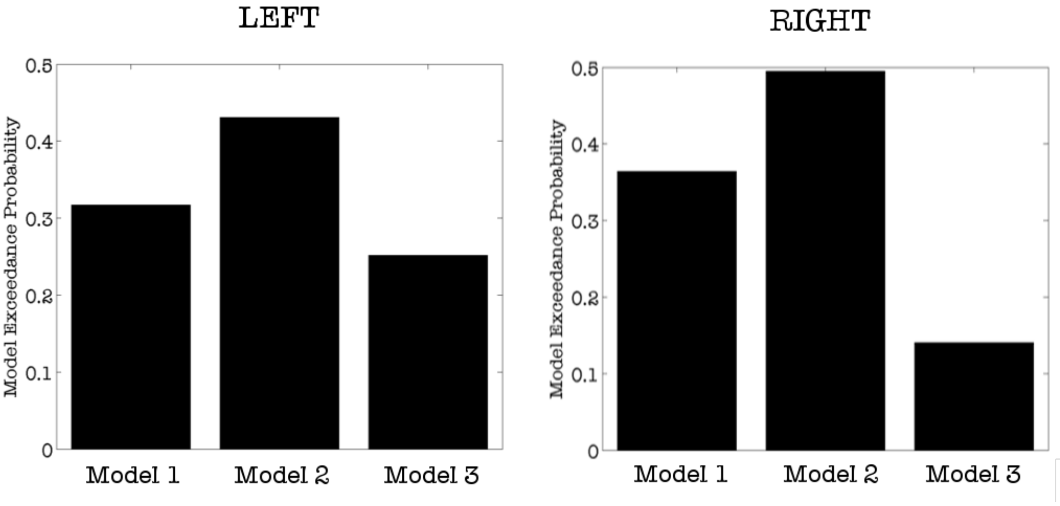

Figure 4.4: The exceedance probability of each model in the model space. Image on the left shows the results of the model testing in the left hemisphere, and the one on the right shows the results of the right hemisphere. In both hemispheres, Model 2 has the highest probability.

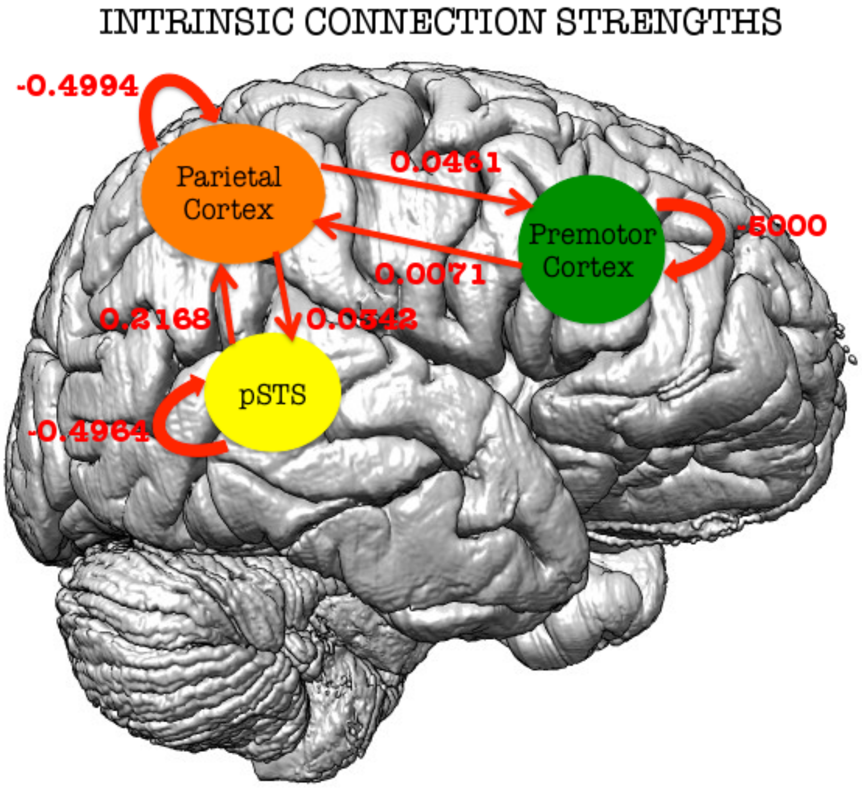

Figure 4.5: Intrinsic connection strengths in the winning model, Model 2 across subjects (only right hemisphere is shown for display purposes). The mean values are also listed in Table 4.2. All connection strengths were significantly different from 0 (with a one-sample t-test, p<0.0001).

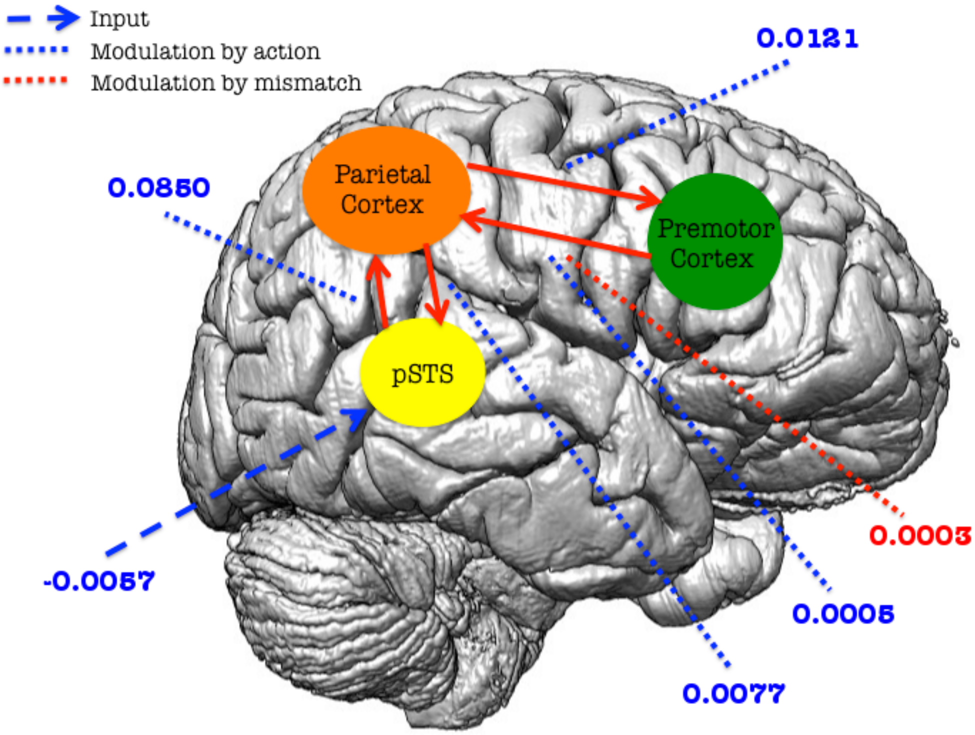

Figure 4.6: Modulatory connection strengths in the winning model, Model 2 across subjects (only right hemisphere is shown for display purposes). The mean values for action modulations for both hemispheres are also listed in Table 4.3. Modulations by actions are shown by the blue dashed lines. Modulation by the mismatch condition is shown by the red dashed line. All connection strengths were significantly different from 0 (with a one-sample t-test, $p<0.0001$).

Table 4.1: The MNI coordinates of central voxels of the ROIs used in the DCM analysis averaged across subjects. pSTS: posterior superior temporal sulcus, STG: superior temporal gyrus, MTG: middle temporal gyrus, IPL: inferior parietal lobule, IFG: inferior frontal gyrus. Values in parenthesis under x, y, z coordinates indicate the standard error of the mean.

	Average MNI coordinates of the central voxels of ROIs in DCM analysis			
	Left hemisphere		Right hemisphere	
Node names in DCM	x y z (Standard error of mean)	Anatomical name/Brodmann Area	x y z (Standard error of mean)	Anatomical name/Brodmann Area
pSTS	-50 -53 7 (1 2 1)	STG/BA 39	53 -44 8 (1 2 1)	MTG/BA 22
Parietal cortex	-37 -42 44 (1 1 1)	IPL/BA7	37 -42 47 (1 1 1)	Subgyral/BA 7
Premotor cortex	-44 6 29 (1 1 1)	IFG/BA 6	47 9 28 (1 2 1)	IFG/BA 44

Table 4.2: The parameters of the intrinsic connectivity that begins with the endogenous activity of actions in the winning model (Model 2) in both hemispheres. The values in the table indicate the mean connection strength across all subjects.

	FROM						
TO		Left hemisphere			Right hemisphere		
		pSTS	Parietal	Premotor	pSTS	Parietal	Premotor
	pSTS	-0.4962	0.0382	-	-0.4964	0.0342	-
	Parietal	0.2016	-0.4994	0.0069	0.2168	-0.4994	0.0071
	Premotor	-	0.0449	-0.5000	-	0.0461	-0.5000

Table 4.3: The parameters of the modulatory activity of actions in the winning model (Model 2) in both hemispheres. The values in the table indicate the mean connection strength across all subjects.

	FROM						
TO		Left hemisphere			Right hemisphere		
		pSTS	Parietal	Premotor	pSTS	Parietal	Premotor
	pSTS	-	0.0095	-	-	0.0077	-
	Parietal	0.0798	-	0.0005	0.0850	-	0.0005
	Premotor	-	0.0121	-	-	0.0121	-

REFERENCES

Friston, K.J., Harrison, L. and Penny, W. (2003). Dynamical causal modeling. Neuroimage, 19(4), 1273-1302.

Friston, K.J. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*, 127–38.

Iacoboni, M. & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience, 7*(12), 942-951.

Kilner, J.M., Vargas, C., Duval, S., Blakemore, S.J., Sirigu, A.(2004). Motor activation prior to observation of a predicted movement. *Nature Neuroscience, 7*:1299–1301.

Kilner, J.M., Friston, K.J., Frith, C.D. (2007a). The mirror-neuron system: a Bayesian perspective. *Neuroreport, 18*:619-623.

Kilner, J.M., Friston, K.J., Frith, C.D. (2007b). Predictive coding: an account of the mirror neuron system. *Cognitive Processing, 8*:159-166.

Luppino, G., Murata, A., Govoni, P., & Matelli, M. (1999). Largely segregated parietofrontal connections linking rostral intraparietal cortex (areas AIP and VIP) and the ventral premotor cortex (areas F5 and F4). *Experimental Brain Research, 128*(1-2), 181-187.

Maranesi, M., Livi, A., Fogassi, L., Rizzolatti, G., Bonini, L. (2014). Mirror neuron activation prior to action observation in a predictable context. *Journal of Neuroscience, 34*:14827–14832

Penny, W.D., Stephan, K.E., Mechelli, A. and Friston, K.J. (2004). Comparing dynamical causal models. *Neuroimage, 22* (3), 1157-72.

Rao, R.P., & Ballard, D.H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*, 79–87.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Reviews of Neuroscience, 27*, 169-192.

Rushworth, M. F. S., Behrens, T.E.J., and Johansen-Berg, H. (2006). Connection patterns distinguish 3 regions of human parietal cortex. *Cerebral Cortex, 16*(10), 1418-1430.

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. *Social Cognitive Affective Neuroscience, 7*(4), 413-422.

Seltzer, B., & Pandya, D. N. (1994). Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: A retrograde tracer study. *The Journal of Comparative Neurology, 343*(3).

Urgen, B.A., Miller. L.E. (2015). Towards an empirically grounded predictive coding account of action understanding. *Journal of Neuroscience*, 35(12), 4789-4791.

CONLCUSION

The overall objective of this dissertation research was to study the functional properties, temporal dynamics and connectivity patterns in the neural system that underlies visual processing of actions, i.e. Action Observation Network in the human brain. More specifically, the focus of this thesis work was whether this system shows specificity for biological agents, particularly for humans. To this end, we used a variety of imaging modalities with excellent spatial and temporal resolution including functional magnetic resonance imaging (fMRI) and electroencephalography (EEG), respectively, together with a variety of analytical techniques such as pattern classification, representational similarity analysis and dynamical causal modeling.

Although studying the basic functional properties and connectivity patterns of the Action Observation Network has improved our understanding about this system, it has also allowed us to address questions regarding human factors in artificial agent design in social robotics and human-robot interaction such as uncanny valley, which is concerned with what kind of robots we should design so that humans can easily accept them as social partners.

In Chapter 1 we aimed to understand the underlying mechanism of the phenomenon uncanny valley (Urgen et al., 2015). Theoretical work and recent empirical work (Saygin et al., 2012a) propose predictive processing as a possible underlying mechanism for uncanny valley but no empirical work to date has directly tested it. Using event-related brain potentials, specifically N400, we provide evidence that supports this hypothesis, which indicate that uncanny valley might be explained by violation of one's expectations about human norms when encountered with very realistic artificial human forms.

In Chapter 2, we investigated the temporal dynamics of the Action Observation Network and its modulation by form and motion of the observed agent using EEG time-frequency analysis, and machine learning (Urgen et al., 2013). Our results indicate that mu oscillations over sensorimotor cortex show a similar pattern during observation of human actions consistent with prior literature. However, we found that it also shows a similar pattern for agents that have varying human-like form and motion. On the other hand, theta oscillations over frontal cortex show sensitivity to the form of the agent observed suggesting that agents that do not look biological enough might results in greater memory processing demands.

In Chapter 3, we investigated the representational properties of brain regions that are known to be involved in action perception using fMRI and a multivariate pattern analysis technique called representational similarity analysis (Kriegeskorte et al., 2008). Our results suggest that different regions that are involved in visual processing of actions have different representational properties. Early visual areas seem to compute the low-level visual properties of the video stimuli, areas in the ventral stream such as FFA seem to be sensitive to the form of the agent, and areas such as LOC and EBA seem to be sensitive to the match between form and motion of the observed agent. The latter results were surprising, as EBA is known to be sensitive to human body forms. These results suggest that EBA might be involved in integration of form and motion rather than pure form processing. On the other hand, core nodes of the Action Observation Network have more similar representational patterns for agents that have biological form compared to the one that has less mechanical form. Among these nodes, pSTS seems to pool information from the visual cortex to compute the identity of the agent, and pass that information to regions in parietal and premotor cortex that code higher-level aspects of actions, consistent with computational models of visual action recognition (Giese and Poggio, 2003).

In Chapter 4, we investigated the flow of information within the Action Observation Network using fMRI and an effective connectivity technique, dynamical causal modeling (DCM) (Friston et al., 2003). Specifically, we investigated the influence of two core nodes, pSTS and premotor cortex over the third node, parietal cortex and how this influence was affected by congruence between the appearance and motion of the observed agent. Our DCM results suggest that the influence on parietal node is likely through a feedback connection from premotor cortex during perception of actions that violate people's predictions, supporting the predictive coding account of action perception (Kilner et al., 2007a; 2007b).

In sum, we had two aims in the set of studies that were presented in this dissertation. First of all, we used artificial agents as experimental tools to probe the human brain and investigate whether it shows specificity for agents that it has evolved with over many generations. On the other hand, we aimed to inform social robotics about what kind of artificial agent we should make. To this end, we used ERPs to study the mechanisms of uncanny valley, cortical EEG oscillations to study the temporal dynamics of the core action observation network, and its modulation by nonhuman agents, fMRI with state-of-the-art pattern analysis to study the representational properties of each region, and DCM to understand the connectivity patterns in this network.

So, what did we learn about the human brain using artificial stimuli? First of all, we show that different areas in the cortical hierarchy in the Action Observation Network have different representational properties consistent with prior computational work. Early stages of visual processing are sensitive to human-like form (Chapter 3). Relatively later stages of processing is sensitive to match between form and motion (Chapter 3), and this sensitivity seems to emerge as early as around 400 ms (Chapter 1), and it is possibly mediated by a top-down influence from premotor cortex to parietal cortex (Chapter 4). On other hand, the temporal dynamics over the sensorimotor system seems to be similar for nonhuman agents as for human agents (Chapter 2).

Next we could ask: What could our basic cognitive neuroscience findings teach to robotics? First of all, it seems like it may be a good idea to make robots that have human-like form as the core nodes of the Action Observation Network show more similar responses to agents that have human-like form (Chapter 3). However, one needs to be aware of uncanny valley, whose underlying mechanism might be expectation violations (Chapter 1). So, it might be a good idea to make robots that do not violate people's expectations. However, what kind of expectations in addition to form-motion match that has been studied in this dissertation, and to what degree is the question to be further investigated in future research. In general, integrating cognitive sciences into the prototyping stage might be useful and fruitful for social robotics.

In conclusion, this thesis on the one hand aims to understand how the human brain responds to agents that it did not evolve with over generations and whether those responses are similar to the ones for humans during perception of actions. On the other hand, it aims to demonstrate that decades of cognitive neuroscience research can allow us to address some fundamental questions in social robotics and human-robot interaction and can guide development of design principles for artificial agents who are increasingly becoming participants in the technology era we live in. In short, it has demonstrated that artificial agents such as robots can allow us to learn about the human brain but the brain science can also inform social robotics (Saygin et al., 2011; Saygin 2012b; Urgen et al., 2013).

REFERENCES

Friston, K.J., Harrison, L. and Penny, W. (2003). Dynamical causal modeling. Neuroimage, *19*(4), 1273-1302.

Giese, M.A. & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*, *4*(3), 179–92.

Kilner, J.M., Friston, K.J., Frith, C.D. (2007a) The mirror-neuron system: a Bayesian perspective. *Neuroreport*, *18*:619-623.

Kilner, J.M., Friston, K.J., Frith, C.D. (2007b) Predictive coding: an account of the mirror neuron system. *Cognitive Processing, 8*:159-166.

Kriegeskorte, N., Mur, M. and Bandettini, P. (2008). Representational Similarity Analysis - Connecting the Branches of Systems Neuroscience. *Frontiers in Systems Neuroscience, 2*(4).

Saygin, A. P., Chaminade, T., Urgen, B.A. and Ishiguro, H. (2011). Cognitive neuroscience and robotics: A mutually beneficial joining of forces. *Robotics: Systems and Science,* Los Angeles, CA.

Saygin, A.P., Chaminade, T., Ishiguro, H., Driver, J. and Frith, C. (2012a). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive Affective Neuroscience, 7*(4): 413-422.

Saygin, A.P. (2012b). What can the Brain Tell us about Interactions with Artificial Agents and Vice Versa? *Workshop on Teleoperated Androids, 34th Annual Conference of the Cognitive Science Society*, Sapporo, Japan.

Urgen, B. A., Plank, M., Ishiguro, H., Poizner, H., & Saygin, A. P. (2013). EEG theta and Mu oscillations during perception of human and robot actions. *Frontiers in Neurorobotics*, *7*, 19.

Urgen, B.A, Li, A.X., Berka, C., Kutas, M., Ishiguro, H., Saygin, A.P. (2015). "Predictive coding and the uncanny valley hypothesis: Evidence from electrical brain activity". *Workshop on Cognition: A Bridge Between Robotics and Interaction* at the *10th ACM/IEEE International Conference on Human-Robot Interaction*, Portland, OR, USA.