The Perception of Humans and Robots:
Uncanny Hills in Parietal Cortex

Ayse Pinar Saygin (saygin@cogsci.ucsd.edu)
Department of Cognitive Science, University of California, San Diego
La Jolla, CA 92093-0515 USA

Thierry Chaminade (tchamina@gmail.com)
Mediterranean Institute for Cognitive Neuroscience, Aix-Marseille University CNRS
Marseille, 13402 France

Hiroshi Ishiguro (ishiguro@ams.eng.osaka-u.ac.jp)
Department of Adaptive Machine Systems, Osaka University
Suita, Osaka, Japan

Abstract

We report on a functional magnetic resonance imaging (fMRI) study of the perception of human and artificial agents. Participants viewed videos of familiar body movements enacted by the android Repliee Q2, the human after whom it was modeled, and the "skinned" version of Q2 revealing its mechanical parts. We used a neural adaptation (repetition suppression) analysis to reveal brain areas sensitive to body movements, and explored whether the identity of the perceived agents modulated these responses. We found significantly higher activity in a distributed network of brain areas for the android, most notably in anterior intraparietal cortex. The responses for the human and the robot with the mechanical appearance resembled each other. We interpret these results within the framework of predictive coding and suggest that the "uncanny valley" phenomenon may have its roots in processing conflicts within the brain's action perception system.

Keywords: action perception; body perception; biological motion; social robotics; artificial agents; neuroimaging; fMRI; uncanny valley

Introduction

In the near future, artificial agents and humanoid robots are expected to be part of our daily lives, not only in entertainment and retail, but also in important domains such healthcare and education (Billard, Robins, Nadel, & Dautenhahn, 2007; Dautenhahn, 2007; Kanda, Ishiguro, Imai, & Ono, 2004). Thus, exploring human factors in interactive robot design and development is crucial (Ishiguro, 2007; MacDorman & Ishiguro, 2006). Conversely, experiments using artificial agents can address questions about the functional properties of mechanisms involved in the perception of others' actions (Blake & Shiffrar, 2007; Rizzolatti & Craighero, 2004). Here, we summarize a neuroimaging study that we performed as part of an interdisciplinary research program that aims to reveal factors that can guide the design of future artificial agents, as well as to improve our understanding of action and body movement perception more generally.

In primates, the perception of body movements is supported by network of lateral superior temporal, inferior parietal and inferior frontal brain areas (Rizzolatti & Craighero, 2004). Here we will refer to this network as the Action Perception System (APS). The frontal and parietal nodes of the system are known to contain mirror neurons, which respond not only when the monkey executes a particular action, but also when it observes another individual perform the action. The existence of a similar system in humans has been suggested by several neuroimaging and lesion studies (e.g., Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995; Grafton, Arbib, Fadiga, & Rizzolatti, 1996; Hari et al., 1998; Iacoboni et al., 1999; Saygin, 2007; Saygin, Wilson, Dronkers, & Bates, 2004).

The neural activity in premotor and parietal regions during action perception is often interpreted within the framework motor resonance, where "an action is understood when its observation causes the motor system of the observer to 'resonate'" (Rizzolatti, Fogassi, & Gallese, 2001). But what are the boundary conditions for this resonance?

There is a small neuroscience literature on the perception of artificial agents, including robots (Chaminade & Hodgins, 2006; MacDorman & Ishiguro, 2006). Unfortunately, the results are not consistent. Some experiments have reported that robot actions affect the observers' own motor processing or the activity of the APS, whereas others have argued that the APS does not respond, or responds weakly if the perceived actor is an artificial agent (Catmur, Walsh, & Heyes, 2007; Chaminade, Hodgins, & Kawato, 2007; Gazzola, Rizzolatti, Wicker, & Keysers, 2007; Kilner, Paulignan, & Blakemore, 2003; Oberman, McCleery, Ramachandran, & Pineda, 2007; Press, Gillmeister, & Heyes, 2007; Tai, Scherfler, Brooks, Sawamoto, & Castiello, 2004). Furthermore, the specific roles of biological appearance or biological motion have not been sufficiently explored in these experiments, but is an area of interest in social robotics, cognitive neuroscience, and vision science (Chaminade, Hodgins, & Kawato, 2007; Cook, Saygin, Swain, & Blakemore, 2009; Kanda, Miyashita, Osada, Haikawa, & Ishiguro, 2008; Minato, Shimada, Itakura, Lee, & Ishiguro, 2006; Oyedele, Hong, & Minor, 2007; Saygin, Wilson, Hagler, Bates, & Sereno, 2004).

On the one hand, it seems reasonable that the closer the match between the observed action and the observers' own sensorimotor representations, the more efficient the simulation will be. In support for this, the APS is modulated by whether the observer can in fact perform the seen movement (Calvo-Merino, Grezes, Glaser, Passingham, & Haggard, 2006; Casile & Giese, 2006). The appearance of the observed agent may be additionally important (Buccino et al., 2004; Chaminade, Hodgins, & Kawato, 2007).

On the other hand, human resemblance is not necessarily always a positive feature in robots. The "uncanny valley" phenomenon points out that as a robot is made more human-like in its appearance, the reaction to it becomes more and more positive and empathetic, until a point is reached at which the robot becomes oddly repulsive (Mori, 1970). The effect is well-known in robotics and animation. For example, the movie *Polar Express* (Warner Bros) was criticized for the characters that viewers found creepy and disturbing. The more recent feature *Avatar* (20th Century Fox) received praise for animations that did not fall into the uncanny valley. Despite such well-known examples, and significant anecdotal evidence, there is little scientific data to characterize the uncanny valley (MacDorman, Green, Ho, & Koch, 2009; Steckenfinger & Ghazanfar, 2009).

The Present Study

This paper briefly describes the approach we took to this topic and summarizes the data from an fMRI repetition suppression study. We performed fMRI as participants viewed video clips of human (H) and robotic agents carrying out recognizable actions. We used Repliee Q2, a humanoid robot developed at Osaka University in collaboration with Kokoro Ltd (Ishiguro et al., 2006). This robot has a very human-like appearance (Figure 1b). In order to achieve this, the robot's face was modeled after an adult Japanese female (Figure 1a). Importantly, Repliee Q2 was videotaped both in its original human-like appearance (the Q2H condition, Figure 1b) and in a modified, more mechanical appearance (the Q2R condition, Figure 1c). In this latter condition, we removed as many of the surface elements as possible in order to reveal the electronics and mechanics underneath. The silicone covering the face and hands could not be removed, so we used a custom mask and gloves to change the appearance of these body parts. The end result was that the robot's appearance became obviously mechanical (e.g., metal arms and joints).

There were three conditions: human (H), robot with human appearance (Q2H) and robot with mechanical appearance (Q2R). However, since the Q2H and Q2R are in fact the same robot, the kinematics are identical for these two conditions. In terms of appearance, H and Q2H are very close to each other, whereas Q2R lies on the mechanical end. In terms of kinematics, H represents truly biological motion and Q2H and Q2R are identical, both with mechanical kinematics.

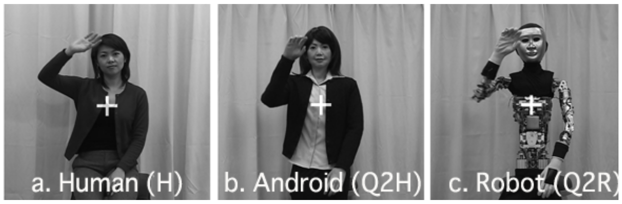

Figure 1. Still frames from the videos used in the experiments depicting the three agents.

The articulators of Repliee Q2 were programmed over several weeks at the Intelligent Robotics Laboratory at Osaka University. The same movements were videotaped in both appearance conditions (Q2R and Q2H). The human (the same female adult to whom Repliee Q2 was designed to resemble) was asked to watch each of Repliee Q2's actions and then perform the same action naturally. All agents were videotaped in the same room and with the same background. A total of 8 actions per actor were used in the experiment, including both transitive (drinking water from a cup, picking up a piece of paper from a table, grasping a tube of hand lotion, wiping a table with a cloth) and intransitive actions (waving hand, nodding affirmatively, shaking head (negative), and introducing self). Video recordings were cut into 2 second long clips, were converted to grayscale, cropped to a uniform size.

20 right handed healthy adults participated. We used a 3T Siemens Allegra scanner at the Wellcome Trust Centre for Neuroimaging in London, UK and a standard T2* weighted gradient echo pulse sequence to obtain functional images (TR=2340 ms, TE=65 ms). 36 slices were acquired at an in-plane resolution of 3 x 3 mm and a through plane resolution of 2 mm and 1 mm gap. Each participant was given exactly the same introduction to the study and the same exposure to the videos prior to scanning since prior knowledge can affect attitudes to artificial agents differentially (Saygin & Cicekli, 2002). Participants were told whether each agent was a human or a robot such that by the time scanning started, they were not uncertain about the identity of the android.

A limitation of previous neuroimaging studies on this topic is that they explored the BOLD fMRI response (Logothetis, 2008). Repetition suppression (henceforth RS, also called fMRI adaptation) is a method applied to fMRI from neurophysiology and refers to the phenomena of reduced neural response to a repeated stimulus compared to the response to a novel stimulus (Grill-Spector & Malach, 2001; Henson & Rugg, 2003; Krekelberg, Boynton, & van Wezel, 2006). RS affects neurons sensitive to the repeated stimulus, so it can be used as a means to explore functional properties of brain areas. In recent years, RS has been applied to the study of action perception (Chong, Cunnington, Williams, Kanwisher, & Mattingley, 2008; Dinstein, Gardner, Jazayeri, & Heeger, 2008; Dinstein, Hasson, Rubin, & Heeger, 2007; Fujii, Hihara, & Iriki, 2008; Hamilton & Grafton, 2006, 2008; Kilner, Neal, Weiskopf, Friston, & Frith, 2009; Lestou, Pollick, & Kourtzi, 2008). This approach was well-suited to our goals as it allows us to test whether neurons in the APS code for biological appearance or biological motion.

Participants watched the action videos in 30 second blocks. There were 12 videos in each block with a 500 ms ISI. Each video was preceded by the same video and the other videos equal number of times and orders were counterbalanced across runs. Each video was preceded by the same video (Repeat) or a different video (Non-repeat). To make sure subjects attended throughout, every 30-seconds, they were presented with a statement about which they made a True/False judgment using a button box (e.g., "I did not see her waving her hand"). The fMRI data were analyzed with SPM5 using standard procedures (http://www.fil.ion.ucl.ac.uk/spm).

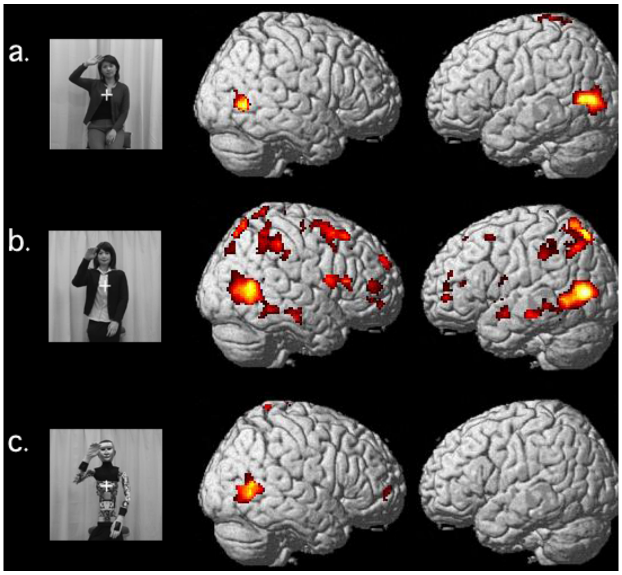

Figure 2. Repetition suppression results for the human (a), Q2H (b), and Q2R(c).

Results

For each agent, we identified regions showing a repetition suppression effect at $p<0.05$ and cluster size of > 30 voxels. The effect of repetition suppression differed between the agents (Figure 2). Posterior temporal cortex showed suppression for all agents, but in the left hemisphere there was significantly less response to Q2R. This area corresponds the Extrastriate Body Area or EBA (Peelen, Wiggett, & Downing, 2006), which responds to the visual perception of the human body.

We otherwise did not find evidence for APS coding for the biological appearance or biological movement of the perceived agents. Instead, in comparison to H and Q2R, a larger network showed suppression for Q2H, despite the use of the same procedures and thresholds. This of course, brings to mind the uncanny valley, except we observed "hills" in the form of increased neural responses rather than valleys. Although we cannot include the details here due to space constraints, a region of interest (ROI) analysis further quantified these results, revealing a significant interaction in the inferior parietal lobule between the agents.

Discussion

We interpret these data within the predictive coding framework, which is based on minimizing prediction error though recurrent interactions among levels of a cortical hierarchy (Bar, 2009; Friston, 2005; Kilner, Friston, & Frith, 2007). During the perception of H and Q2R, where there is no mismatch between the appearance and the movement of the agent. For Q2H on the other hand, there is a human-like appearance that leads to a conflict when this information is integrated with the movement kinematics of the agent. This will lead to the generation of a prediction error, which is propagated in the network until the errors of each node are minimized. It is possible to measure prediction errors using neuroimaging (Friston, 2010). It is not possible from the current data to know the exact source and time course of error propagation, but it is clear that the cortical network is engaged strongly during the perception of Q2R compared with the agents that lead to less prediction error. The effect is largest in parietal cortex, which is the node of the network that links the posterior, visual components of the APS and the frontal, motor components (Matelli & Luppino, 2001; Seltzer & Pandya, 1994).

The present study is only a beginning. This framework provides hypotheses that we are testing in new studies. We are now utilizing animation to modulate the appearance and movement parameters more precisely (although this may lead to decrease in presence (Sanchez-Vives & Slater, 2005), whose importance in modulating APS is currently not known). We also need to use other neuroimaging and psychological methods in addition to, or in conjunction with fMRI to study the temporal dynamics of action processing.

With brief exposure times, Repliee Q2 can be mistaken for a human being, but longer exposure usually triggers the feeling of repulsion or discomfort characteristic of the uncanny valley (Ishiguro, 2006). While we did not explicitly assess the uncanny valley in this study, our results suggest an intriguing relationship between the APS and this phenomenon. We are currently exploring this in more sophisticated analyses as well as with new experiments.

In summary, we found that a robot with very humanlike appearance can cause differential responses compared with the same robot with a mechanical appearance, or with a human being that maximally resembles the robot. These differences were found in a network of brain areas, but most prominently in inferior parietal cortex, which connects the posterior areas involved in the visual perception of actions and biological motion to premotor areas in frontal cortex. We propose these "hills" in the brain activity reflect the prediction error that is generated as the brain processes these stimuli. We suggest that the uncanny valley may arise from processing conflicts in the APS, and can be investigated using fMRI.

Acknowledgments

This research was supported by an innovative research grant to A.P. Saygin from the Kavli Institute for Brain and Mind (UCSD). Additional support was contributed by the European Commission and by the Wellcome Trust. We are grateful to members of the Intelligent Robotics Laboratory for their help in creating the experimental stimuli and to Jon Driver, Chris Frith, James Kilner and members of the Wellcome Trust Centre for Neuroimaging for their support of the fMRI study. We also appreciate discussions with Karl MacDorman, Takashi Minato, Javier Movellan, and Marty Sereno in the early stages of this project.

References

Bar, M. (2009). Predictions: a universal principle in the operation of the human brain. Introduction. *Philos Trans R Soc Lond B Biol Sci, 364*(1521), 1181-1182.

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*(1), 37-49.

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annu Rev Psychol, 58*, 47-73.

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. *J Cogn Neurosci, 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. *Curr Biol, 16*(19), 1905-1910.

Casile, A., & Giese, M. A. (2006). Nonvisual motor training influences biological motion perception. *Curr Biol, 16*(1), 69-74.

Catmur, C., Walsh, V., & Heyes, C. (2007). Sensorimotor learning configures the human mirror system. *Curr Biol, 17*(17), 1527-1531.

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

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

Chong, T. T., Cunnington, R., Williams, M. A., Kanwisher, N., & Mattingley, J. B. (2008). fMRI adaptation reveals mirror neurons in human inferior parietal cortex. *Curr Biol, 18*(20), 1576-1580.

Cook, J., Saygin, A. P., Swain, R., & Blakemore, S. J. (2009). Reduced sensitivity to minimum-jerk biological motion in autism spectrum conditions. *Neuropsychologia*.

Dautenhahn, K. (2007). Socially intelligent robots: dimensions of human-robot interaction. *Philos Trans R Soc Lond B Biol Sci, 362*(1480), 679-704.

Dinstein, I., Gardner, J. L., Jazayeri, M., & Heeger, D. J. (2008). Executed and observed movements have different distributed representations in human aIPS. *J Neurosci, 28*(44), 11231-11239.

Dinstein, I., Hasson, U., Rubin, N., & Heeger, D. J. (2007). Brain areas selective for both observed and executed movements. *J Neurophysiol, 98*(3), 1415-1427.

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

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions B, 360*(1456), 815.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat Rev Neurosci, 11*(2), 127-138.

Fujii, N., Hihara, S., & Iriki, A. (2008). Social cognition in premotor and parietal cortex. *Soc Neurosci, 3*(3-4), 250-260.

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., Arbib, M. A., Fadiga, L., & Rizzolatti, G. (1996). Localization of grasp representations in humans by positron emission tomography. 2. Observation compared with imagination. *Exp Brain Res, 112*(1), 103-111.

Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol (Amst), 107*(1-3), 293-321.

Hamilton, A. F., & Grafton, S. T. (2006). Goal representation in human anterior intraparietal sulcus. *J Neurosci, 26*(4), 1133-1137.

Hamilton, A. F., & Grafton, S. T. (2008). Action outcomes are represented in human inferior frontoparietal cortex. *Cereb Cortex, 18*(5), 1160-1168.

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. *Proc Natl Acad Sci U S A, 95*(25), 15061-15065.

Henson, R. N., & Rugg, M. D. (2003). Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia, 41*(3), 263-270.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science, 286*(5449), 2526-2528.

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

Ishiguro, H. (2007). Projects and Vision in Robotics. *Lecture Notes in Computer Science, 4314*, 451.

Ishiguro, H., Asada, M., Shapiro, S. C., Thielscher, M., Breazeal, C., Mataric, M. J., et al. (2006). Human-Inspired Robots. *IEEE Intelligent Systems, 21*(4), 74-85.

Kanda, T., Ishiguro, H., Imai, M., & Ono, T. (2004). Development and evaluation of interactive humanoid robots. *Proceedings of the IEEE, 92*(11), 1839-1850.

Kanda, T., Miyashita, T., Osada, T., Haikawa, Y., & Ishiguro, H. (2008). Analysis of humanoid appearances in human-robot interaction. *IEEE Transactions on Robotics, 24*(3), 725-735.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). The mirror-neuron system: a Bayesian perspective. *Neuroreport, 18*(6), 619-623.

Kilner, J. M., Neal, A., Weiskopf, N., Friston, K. J., & Frith, C. D. (2009). Evidence of mirror neurons in human inferior frontal gyrus. *J Neurosci, 29*(32), 10153-10159.

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

Krekelberg, B., Boynton, G. M., & van Wezel, R. J. (2006). Adaptation: from single cells to BOLD signals. *Trends Neurosci, 29*(5), 250-256.

Lestou, V., Pollick, F. E., & Kourtzi, Z. (2008). Neural substrates for action understanding at different description levels in the human brain. *Journal of Cognitive Neuroscience, 20*(2), 324-341.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature, 453*(7197), 869-878.

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

Matelli, M., & Luppino, G. (2001). Parietofrontal circuits for action and space perception in the macaque monkey. *Neuroimage, 14*(1 Pt 2), S27-32.

Minato, T., Shimada, M., Itakura, S., Lee, K., & Ishiguro, H. (2006). Evaluating the human likeness of an android by comparing gaze behaviors elicited by the android and a person. *Advanced Robotics, 20*(10), 1147.

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

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.

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

Peelen, M. V., Wiggett, A. J., & Downing, P. E. (2006). Patterns of fMRI activity dissociate overlapping functional brain areas that respond to biological motion. *Neuron, 49*(6), 815-822.

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

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annu Rev Neurosci, 27*, 169-192.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat Rev Neurosci, 2*(9), 661-670.

Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nat Rev Neurosci, 6*(4), 332-339.

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

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

Saygin, A. P., Wilson, S. M., Dronkers, N. F., & Bates, E. (2004). Action comprehension in aphasia: linguistic and non-linguistic deficits and their lesion correlates. *Neuropsychologia, 42*(13), 1788-1804.

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. *J Neurosci, 24*(27), 6181-6188.

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).

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

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