What can the Brain Tell us about Interactions with Artificial Agents and Vice Versa?

Ayse Pinar Saygin (saygin@cogsci.ucsd.edu)
Department of Cognitive Science, 9500 Gilman Drive
University of California, San Diego
La Jolla, CA 92093-0515 USA

Abstract

No longer encountered only in science fiction, artificial agents such as humanoid robots and interactive animated characters are rapidly becoming participants in many aspects of social and cultural life. Artificial agents have a range of biomedical, educational and entertainment applications. In particular, they can enable telepresence, opening a range of new possibilities for human interaction. For these technologies to succeed however, we need to understand human factors guiding our interactions with these agents. In our research we use methods from cognitive neuroscience and neuroimaging to explore how humans perceive, respond to, and interact with others, including artificial agents. Not only can we inform the design of new agents by studying human brain responses in interactions with artificial agents, but studies with artificial agents can improve our understanding of how the human brain enables some of our most important skills such as action understanding, social cognition, empathy, and communication. We suggest interdisciplinary collaboration is the most fruitful way to proceed in advancing robotics and animation on one hand, and cognitive science and neuroscience on the other.

Keywords: action perception; uncanny valley; mirror neurons; biological motion

Introduction

With advances in technology, artificial agents such as robots are quickly becoming parts of our daily lives (Coradeschi et al., 2006; Ishiguro & Nishio, 2007). These technologies can enable telepresence, opening up new possibilities in human interaction that can reduce costs and travel (and associated carbon emissions), as well as increase diversity of participation. Thus, research on how humans perceive, respond to and interact with these agents is increasingly important (MacDorman & Kahn Jr, 2007; Sanchez-Vives & Slater, 2005; Saygin, Chaminade, Urgen, & Ishiguro, 2011). In particular, neuroscience and psychology research exploring human robot interaction (HRI) and telepresence can make valuable contributions to the development of future applications (Chaminade & Cheng, 2009; Chaminade & Hodgins, 2006; Saygin et al., 2011). An interdisciplinary perspective on human-agent interaction is especially important, since this field will impact issues of public concern in the near future, for example in domains such as education and healthcare (e.g., Billard, Robins, Nadel, & Dautenhahn, 2007; Kanda, Ishiguro, Imai, & Ono, 2004; Mataric, Tapus, Winstein, & Eriksson, 2009).

Conversely, experiments on the perception of artificial agents and telepresence can help advance neuroscience, since they can help us explore the functional properties of brain areas that subserve social cognition (e.g., Chaminade et al., 2010; Cross et al., 2011; Gazzola, Rizzolatti, Wicker, & Keysers, 2007; Saygin, Chaminade, Ishiguro, Driver, & Frith, 2012). Using artificial agents and telepresence we can control stimulus properties precisely, or create entities or environtments that violate physical realities of the world. Such manipuations can allow us to test whether particular neural systems or perceptual processes are selective or sensitive to natural (biological) stimuli or might also generalize to non-biological (artificial) stimuli.

The goal of our research program is to both improve our understanding of how the human brain enables social cognition, and to help engineers and designers in developing interactive agents that are well-suited to their application domains, as well as to the brains of their creators. In this paper, I will give an example of a neuroimaging study in which we have used artificial agents (humanoid robots) to study the human brain. Such interdisciplinary work that can allow us to answer questions about both artificial agents and about the brain are important as we face a future that includes interactions with such agents and telepresence.

Action Perception

In primates, the perception of body movements and actions is supported by network of lateral superior temporal, inferior parietal and inferior frontal brain areas. Here, we refer to this network as the action perception system, or APS (Fig. 1). Two of the areas within the APS, (PMC and IPL) contain mirror neurons in the macaque brain (Rizzolatti & Craighero, 2004). Mirror neurons respond not only when a monkey executes a particular action, but also when it observes another individual perform the action. 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. It is thought that a similar system underlies action perception in the human brain (e.g., Grafton, 2009; Iacoboni & Dapretto, 2006; Saygin, 2007; Saygin, Wilson, Hagler,

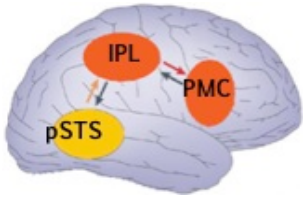

Figure 1: Schematic of the Action Perception System (APS): Superior Temporal Sulcus (pSTS), Inferior Parietal Lobule (IPL), Premotor Cortex (PMC). Adapted from Iacoboni & Dapretto, 2006.

Bates, & Sereno, 2004). Some researchers have argued that in addition to subserving action processing, the APS helps in linking "self" and "other", and thus may constitute a basis for social cognition (Rizzolatti & Craighero, 2004).

The finding that the visual perception of another entity automatically engages the observers' own motor system indicates that at some levels of the nervous system, simply seeing another agent automatically engages interaction.

The APS has received intense interest from neuroscientists in the last decade and a half, and we can now use the accumulated knowledge in this field to study how the human brain supports interactions with artificial agents and telepresence. Conversely research on artificial agent perception and telepresence can help research on the human brain by allowing us to test functional properties of the APS and other brain areas.

Due to the presence of mirror neurons, the neural activity in PMC and IPL regions during action perception is often interpreted within the framework of "simulation": A visually perceived body movement 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, Fogassi, & Gallese, 2001). But what are the boundary conditions for 'resonance'? What kinds of agents or actions lead to the simulation process? Is human-like appearance important? Is human-like motion?

On the one hand, we might expect the closer the match between observed action and 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 (e.g., Calvo-Merino, Grezes, Glaser, Passingham, & Haggard, 2006). The appearance of the observed agent may also be important (e.g., Chaminade, Hodgins, & Kawato, 2007).

On the other hand, human resemblance is not necessarily always a positive feature in artificial agents. The "uncanny valley" 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). While this phenomenon is well known to roboticists and animators, there is only a small (but growing) body of experimental evidence in favor of or against it (e.g., Cheetham, Suter, & Jancke, 2011; Ho, MacDorman, & Dwi Pramono, 2008; Lewkowicz & Ghazanfar, 2012; MacDorman & Ishiguro, 2006; Saygin et al., 2012; Seyama & Nagayama, 2007; Steckenfinger & Ghazanfar, 2009; Thompson, Trafton, & McKnight, 2011; Tinwell, Grimshaw, Nabi, & Williams, 2011). The uncanny valley not only constitutes a practical challenge for robotics and telepresence, but also is a puzzling phenomenon to study from a perceptual and cognitive standpoint.

Robots can have nonbiological 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 the human Action Perception System (APS)? Robots can allow us to ask such questions and to test whether particular brain areas are selective or sensitive to the presence of a human, or an agent with a humanlike form, or respond regardless of the agent performing the action.

Neuroimaging Study: Perception of Robot and Android Actions

There is a small neuroscience literature on the perception of artificial agents, including robots (e.g., Gazzola et al., 2007; Oberman, McCleery, Ramachandran, & Pineda, 2007; Tai, Scherfler, Brooks, Sawamoto, & Castiello, 2004). Unfortunately, the results are highly inconsistent. Furthermore, many studies had used toy robots or very rudimentary industrial robot arms, so the results were not informative regarding state-of-the-art humanoid robots or telepresence. Furthermore, the roles of humanlike appearance or motion were not explored in previous work. We used neuroimaging (functional Magnetic Resonance Imaging (fMRI)) along with a method called Repetition Suppression (RS) to overcome limitations of previous work, and studied this question with well-controlled stimuli developed in by an interdisciplinary team (Saygin, Chaminade, & Ishiguro, 2010; Saygin et al., 2012).

We performed fMRI as participants viewed video clips of human and robotic agents carrying out recognizable actions. fMRI is a powerful method that allows imaging the activity of the live human brain non-invasively and has revolutionized neuroscience, though as with any method, there are limitations (e.g., no ferromagnetic materials, limited interactivity).

We used Repliee Q2, an android developed at Osaka University in collaboration with Kokoro Ltd (Ishiguro, 2006; Ishiguro & Nishio, 2007). Repliee Q2 has a very human-like appearance (Fig. 2, Android (A)); the robot's face was modeled after an adult Japanese female who also participated in our stimulus development (Fig. 2, Human (H)). Repliee Q2 can make facial expressions, as well as eye, head, upper limb, and torso movements. It has 42 degrees of freedom (d.o.f.) in its movements, with 16 d.o.f. in the head. With very brief exposure times, Repliee Q2 is often mistaken for a human being, but more prolonged exposure and interaction can lead to an uncanny valley experience (Ishiguro, 2006).

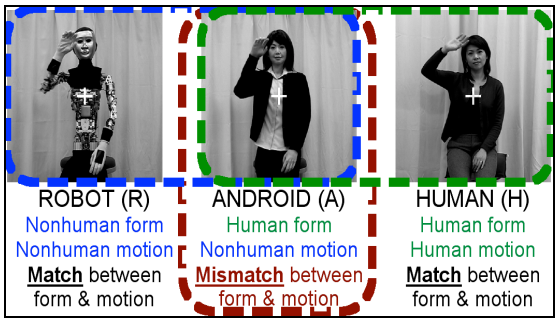

Figure 2: Stills from the videos depicting the three agents (R, A, H) and the experimental conditions (form and motion) they represent.

Repliee Q2 was videotaped both in its original human-like appearance (A) and in a modified, more mechanical appearance (Fig. 2, Robot (R)). For this, 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 cover these areas. The end result was that the robot's appearance became mechanical and nonhuman. However, since the A and R are in fact the same robot, the motion dynamics and kinematics are the same for these two conditions.

There were thus three agents: human (H), robot with human form (A), and robot with nonhuman form (R). H and A are very close to each other in form, both with humanlike form, whereas R has nonhuman form. In terms of the movement, H represents truly biological motion and A and R are identical, both with mechanical kinematics. Using fMRI and RS, we explored whether the human brain would display specialization for human form (similar responses for A and H, and different for R) or motion (similar responses for R and A, and differential responses for H). Another possibility was for RS responses not to reflect biological form or motion per se, but instead pattern with the uncanny valley. In this scenario, responses to H and R would be similar to each other, even though these two agents are divergent from each other in both form and movement.

The articulators of Repliee Q2 were programmed over several weeks at Osaka University. The same movements were videotaped in both appearance conditions (R and A). The human, the same female adult to whom Repliee Q2 was designed to resemble, was asked to perform the same actions as she naturally would. 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 (e.g., drinking water from a cup, waving hand). 20 adults participated in the fMRI experiment. Participants had no experience working with robots. Each 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). Before the experiment, subjects were told that they would see short video clips of actions by a person, or by two robots with different appearances and were shown all the movies in the experiment. By the time scanning started, participants were not uncertain about the robotic identity of the android.

Scanning was conducted at the Wellcome Trust Centre for Neuroimaging, in London, UK using a 3T Siemens Allegra scanner and a standard T2* weighted gradient echo pulse sequence. During fMRI, subjects viewed the stimuli projected on a screen in the back of the scanner bore through a mirror placed inside the head coil. There were blocks of 12 videos, each preceded by the same video (Repeat) or a different video (Non-repeat), which allowed us to compute the RS contrast (Non-repeat > Repeat). Every 30-seconds, they were presented with a statement about which they would have to make a True/False judgment (e.g., "I did not see her wiping the table"). Since the statements could refer to any video, subjects had to be attentive throughout the block. Data were analyzed with SPM software (http://www.fil.ion.ucl.ac.uk/spm).

RS differed considerably between the agents (Fig. 3). All agents showed RS in temporal cortex near the pSTS. For A, extensive RS was found in additional regions of temporal, parietal and frontal cortex (Fig. 3b).

In the left hemisphere, lateral temporal cortex responded to H and A, but not to R. The specific location of this activation corresponds to extrastriate body area (EBA), a region that responds strongly during the visual perception of the body and body parts (Peelen, Wiggett, & Downing, 2006). Our data showed that robotic appearance can weaken the RS response in the EBA.

Aside from the EBA, we did not find evidence selective coding for human form or motion. Instead, for A, whose form is humanlike, but its motion mechanical, increased responses were found in a network of cortical areas. This was most pronounced (and statistically significant) in the IPL, one of the nodes of the APS (Fig. 3b, circled areas).

But why would there be an area of the brain highly selective for androids? This response pattern brings to mind the uncanny valley – except, rather than valleys, we measured "hills" in the neural responses, in the form of increased RS. A framework within which to interpret these data is the predictive coding account of cortical computation (Friston & Kiebel, 2009; Friston, Mattout, & Kilner, 2011; Kilner, Friston, & Frith, 2007). Predictive coding is based on minimizing prediction error among the levels of a cortical hierarchy (e.g. the APS). More specifically, during the perception of H and R, there is no conflict between form and motion of the agent. H appears human and moves like a human. R appears mechanical and moves mechanically. For A on the other hand the agent's form is humanlike, which may result in a conflict when the brain attempts to process

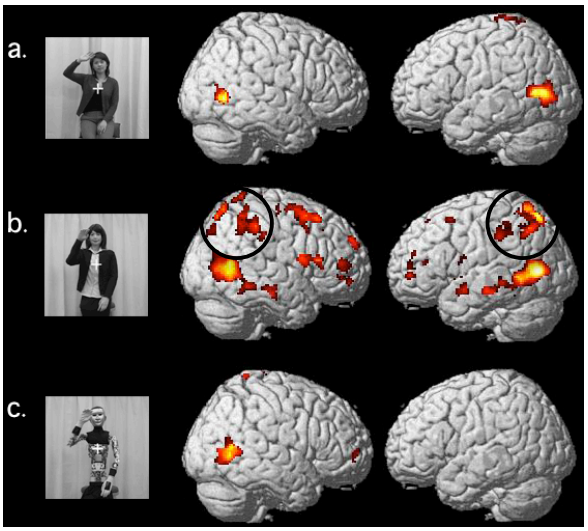

Figure 3. Repetition suppression (RS) results for the Human (a), Android (b), and Robot (c). (Non-repeat > Repeat at t>=8.86, p<0.05 with False Discovery Rate (FDR) correction for multiple comparisons, cluster size of at least 30 voxels). Adapted from Saygin et al., 2012.

and integrate the movement of the agent with its form. This conflict leads to the generation of a prediction error, which is propagated in the network until the predictions of each node are minimized. During this process, we can measure the prediction error in the fMRI responses. It is not possible from the current data to know the exact neural sources, the directionality, and the time course of error propagation, but it is clear that the cortical network is engaged more strongly during the perception of A compared with the agents that lead to less prediction error (R and H). Furthermore, the effect is largest in parietal cortex, the node of the network that links the posterior, visual components of the APS and the frontal, motor components (Matelli & Luppino, 2001).

In summary, in this interdisciplinary study, we found that a robot with highly humanlike form is processed differentially compared with a robot with a mechanical form, or with an actual human. These differences are found in a network of brain areas, most prominently in parietal cortex (Saygin et al., 2012). We propose these "hills" in the brain activity reflect the prediction error that is propagated in the system. The uncanny valley may thus arise from processing conflicts in the APS, and the resultant error signals, which can in turn be measured using fMRI.

The study described above constitute only a beginning. In future work, we can utilize animation in order to modulate form and motion parameters more precisely (although this is likely to lead to a decrease in presence (Sanchez-Vives & Slater, 2005)). We will also use other neuroimaging and psychological methods in addition to, or in conjunction with fMRI. More time-resolved behavioral and neuroimaging methods are also important to study the temporal dynamics of action processing (Saygin & Stadler, 2012; Urgen, Plank, Ishiguro, Poizner, & Saygin, 2012).

Discussion

Using cognitive neuroscience, we have been able to suggest an interpretation for the classic anecdotal reports of the uncanny valley hypothesis. While our experiment was not designed to explain the uncanny valley, the results suggest an intriguing link between the phenomenon, and brain responses in the APS. As shown in Figure 2, the android condition features a mismatch between form and motion. In a predictive coding, the android is not predictable: an agent with that form (human) would typically not move mechanically as Repliee Q2 does. When the nervous system is presented with this unexpected combination, a propagation of prediction error may occur in the APS. We suggest this framework may contribute to an explanation for the uncanny valley and future experiments will test this hypothesis.

Using robotics, we were able to answer questions regarding the neural basis of action perception. We were able to test functional properties of human action perception system (APS), helping shed light on how our brains enable social cognition.

Collaboration between cognitive neuroscience and robotics and telepresence research can be a win-win for both sides. Understanding both the computational and the human side of human-agent interaction is necessary for developing successful assistive artificial agents and telepresence systems.

Acknowledgments

This research was supported by the Kavli Institute for Brain and Mind, California Institute of Telecommunications and Information Technology (Calit2), NSF (CAREER-1151805), Marie Curie Intra-European Fellowship, and the Wellcome Trust. I would like to thank our coauthors, collaborators and colleagues who have commented on this work: Thierry Chaminade, Jon Driver, Chris Frith, Antonia Hamilton, Hiroshi Ishiguro, Javier Movellan, Takashi Minato, Marty Sereno; and the coauthors on our EEG and MEG studies: James Kilner, Marta Kutas, Howard Poizner, Markus Plank and Burcu Urgen.

References

Billard, A., Robins, B., Nadel, J., & Dautenhahn, K. (2007). Building Robota, a mini-humanoid robot for the rehabilitation of children with autism. *Assistive Technologies, 19*(1), 37-49.

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*(19), 1905-1910.

Chaminade, T., & Cheng, G. (2009). Social cognitive neuroscience and humanoid robotics. *Journal of Physiology Paris, 103*(3-5), 286-295.

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.

Chaminade, T., Zecca, M., Blakemore, S. J., Takanishi, A., Frith, C. D., Micera, S., . . . Umilta, M. A. (2010). Brain response to a humanoid robot in areas implicated in the perception of human emotional gestures. *PLoS ONE, 5*(7), e11577.

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.

Coradeschi, S., Ishiguro, H., Asada, M., Shapiro, S. C., Thielscher, M., Breazeal, C., . . . Ishida, H. (2006). Human-inspired robots *IEEE Intelligent Systems, 21*(4), 74-85.

Cross, E. S., Liepelt, R., de, C. H. A. F., Parkinson, J., Ramsey, R., Stadler, W., & Prinz, W. (2011). Robotic movement preferentially engages the action observation network. *Human Brain Mapping*.

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society London B, 364*(1521), 1211-1221.

Friston, K., Mattout, J., & Kilner, J. (2011). Action understanding and active inference. *Bioligical Cybernetics, 104*(1-2), 137-160.

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. (2009). Embodied cognition and the simulation of action to understand others. *Annual NY Academy of Sciences, 1156*, 97-117.

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*. 3rd ACM/IEEE international conference on Human robot interaction, Amsterdam, Netherlands.

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.

Ishiguro, H., & Nishio, S. (2007). Building artificial humans to understand humans. *Journal of Artifical Organs, 10*(3), 133-142.

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

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

Lewkowicz, D. J., & Ghazanfar, A. A. (2012). The development of the uncanny valley in infants. *Developmental Psychobiology, 54*(2), 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., & Kahn Jr, P. H. (2007). Introduction to the special issue on psychological benchmarks of human-robot interaction. *Interaction Studies, 8*(3), 359-362.

Mataric, M., Tapus, A., Winstein, C., & Eriksson, J. (2009). Socially assistive robotics for stroke and mild TBI rehabilitation. *Studies in Health Technology Informatics, 145*, 249-262.

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

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.

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.

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

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

Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience, 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., Chaminade, T., & Ishiguro, H. (2010). *The perception of humans and robots: Uncanny hills in parietal cortex*. Proceedings of the 32nd Annual Conference of the Cognitive Science Society, Portland, OR.

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

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

Saygin, A. P., & Stadler, W. (2012). The role of appearance and motion in action prediction. *Psychological Research, 76*(4), 388-394.

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.

Seyama, J., & 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., & 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. *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*, 695–705.

Tinwell, A., Grimshaw, M., Nabi, D. A., & Williams, A. (2011). Facial expression of emotion and perception of the Uncanny Valley in virtual characters. *Computers in Human Behavior, 21*, 741–749.

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*. 34rd Annual Conference of the Cognitive Science Society, Sapporo, Japan.