Social Cognitive and Affective Neuroscience Advance Access published April 22, 2011

doi:10.1093/scan/nsr025 SCAN (2011) 1 of 10

The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions

Ayse Pinar Saygin,[1,2] Thierry Chaminade,[2,3] Hiroshi Ishiguro,[4,5] Jon Driver,[2] and Chris Frith[2,6]

[1]Department of Cognitive Science and Neurosciences Program, University of California, San Diego, La Jolla, CA, USA, [2]Wellcome Trust Centre for Neuroimaging, University College London, London, UK, [3]Mediterranean Institute for Cognitive Neuroscience (INCM), CNRS - Aix-Marseille Université, Marseille, France, [4]Department of Systems Innovation, Osaka University, Osaka, Japan, [5]Intelligent Robotics and Communication Laboratories, ATR, Keihanna Science City, Kyoto, Japan, and [6]University of Aarhus, Denmark

Using functional magnetic resonance imaging (fMRI) repetition suppression, we explored the selectivity of the human action perception system (APS), which consists of temporal, parietal and frontal areas, for the appearance and/or motion of the perceived agent. Participants watched body movements of a human (biological appearance and movement), a robot (mechanical appearance and movement) or an android (biological appearance, mechanical movement). With the exception of extrastriate body area, which showed more suppression for human like appearance, the APS was not selective for appearance or motion *per se*. Instead, distinctive responses were found to the mismatch between appearance and motion: whereas suppression effects for the human and robot were similar to each other, they were stronger for the android, notably in bilateral anterior intraparietal sulcus, a key node in the APS. These results could reflect increased prediction error as the brain negotiates an agent that appears human, but does not move biologically, and help explain the 'uncanny valley' phenomenon.

Keywords: functional magnetic resonance imaging (fMRI); repetition suppression; action perception; predictive coding; temporal cortex; anterior intraparietal sulcus; mirror neuron system; extrastriate body area

INTRODUCTION

Understanding others' movements and actions is important for many tasks of ecological significance, such as hunting prey, avoiding predators, communication and social interaction. How humans and other animals achieve this has long been of interest in psychology and neuroscience (Blake and Shiffrar, 2007). In primates, perception of body movements is thought to be supported by a network including lateral superior temporal, inferior parietal and inferior frontal brain areas. Neuroimaging studies have shown responses in these areas during observation of actions; neuropsychological patient and transcranial magnetic stimulation (TMS) studies have shown that damage or disruption of these areas can affect action processing (Saygin *et al.*, 2004a; Pobric and Hamilton, 2006; Grafton and Hamilton, 2007 Saygin, 2007; Candidi *et al.*, 2008). In non-human primates, at least two of these regions have been reported to contain 'mirror neurons', which fire during the execution as well as the observation of specific movements (Rizzolatti and Craighero, 2004). Hence, this network is sometimes referred to as the mirror-neuron system (MNS). The exact relationship between mirror neurons and brain areas that support action perception in the human brain remains a topic of debate (e.g. Dinstein *et al.*, 2007; Chong *et al.*, 2008; Kilner *et al.*, 2009; Mukamel *et al.*, 2010). Accordingly, we will refer to the brain areas most commonly discussed in relation to action perception (i.e. lateral temporal, inferior frontal/ventral premotor and anterior intraparietal cortex) more neutrally as the *Action Perception System* (APS), although of course action perception may involve other parts of the brain as well.

Observed neural activity in the APS is often interpreted within the framework of motor resonance, whereby '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 this resonance? How similar do the actors have to be with respect to the observer to engage resonance?

On the one hand, it has been argued that closer the match between the observed action and the observers' own sensorimotor representations, the stronger the resonance should be. In support for this, there are links between activity within the APS and whether the observer can perform the seen movement (e.g. Calvo-Merino *et al.*, 2006; Cross *et al.*, 2006; Candidi *et al.*, 2008). The appearance of the observed agent

Received 22 July 2011; Accepted 16 March 2011

The authors thank Patrick Haggard, Antonia Hamilton, James Kilner, Takashi Minato, Javier Movellan, Marty Sereno, Osaka University Intelligent Robotics Laboratory and the Wellcome Trust Centre for Neuroimaging. This research was funded by an innovative research grant to APS from the Kavli Institute for Brain and Mind, University of California, San Diego. APS was additionally supported by California Institute for Telecommunication and Information Technology (Calit2); CF by Danish National Research Foundation; HI by Japan Society for the Promotion of Science; JD by The Royal Society. Repliee Q2 was developed in collaboration with Kokoro Inc.

Correspondence should be addressed to Ayse Pinar Saygin, 9500 Gilman Drive, Department of Cognitive Science, MC 0515, La Jolla, CA 92093-0515, USA. Email: saygin@cogsci.ucsd.edu

may also be important (Buccino *et al.*, 2004; Chaminade *et al.*, 2007). On the other hand, responses in the APS can appear surprisingly insensitive to the surface properties of the viewed action stimuli. For example, in the human brain, parts of the APS respond to actions and body movements of simple animations (Pelphrey *et al.*, 2003) or to point-light displays (Saygin *et al.*, 2004b). Indeed some researchers have suggested that the system is sensitive to the action's meaning, but is relatively insensitive to the surface properties of the sensory signals transmitting this information (Craighero *et al.*, 2007).

While humans have long been preoccupied with the theme of creating other entities in their likeness (e.g. dolls, marionettes, stories like the Golem, Frankenstein), with technological advances, artificial agents such as humanoid robots and 3D animated characters are becoming more and more commonly encountered in daily life (Coradeschi *et al.*, 2006). Artificial agents can also provide scientists with unique opportunities to test theories of human perception and cognition. For example, robots can have appearance or movement kinematics that are not biological, but can nevertheless be perceived as carrying out recognizable actions. They can thus be used to study the functional properties of the APS, such as whether the network is tuned selectively to human-like appearance, or biological motion.

There is a small neuroscience literature on the perception of actions of artificial agents, including robots. Unfortunately, the results are not consistent to date. Some studies have reported that artificial agents' actions apparently affect the observers' own motor processing, or activity within the APS, whereas others have argued that the APS either does not respond, or responds weakly if the perceived actor is not human (e.g. Kilner *et al.*, 2003; Tai *et al.*, 2004; Chaminade and Hodgins, 2006; Catmur *et al.*, 2007; Chaminade *et al.*, 2007; Gazzola *et al.*, 2007; Oberman *et al.*, 2007; Press *et al.*, 2007). The specific roles of biological appearance *vs* biological motion have not been sufficiently explored or separated in previous studies, even though this is a topic of increasing interest in robotics, neuroscience and vision science (MacDorman and Ishiguro, 2006; Chaminade *et al.*, 2007, 2010; Kanda *et al.*, 2008; Jastorff and Orban, 2009; Saygin *et al.*, 2010).

In the present study, our stimuli and experimental design focused on whether the seen agent had biological (human-like) *appearance* and also whether the agent's body *movements* were biological, plus whether their appearance and movements matched. We also manipulated repetition of successive actions, as explained below.

While our interest was focused on the APS, it was not limited to these regions alone. For example, the involvement of form processing in biological motion perception has also been supported (e.g. Lange and Lappe, 2006). Our methods allowed us to explore regions of the brain involved in body movement perception without limiting our focus to the nodes of the APS.

A novel aspect of this study was that we used a recently developed, state-of-the-art android,[1] Repliee Q2. This was important for several reasons. First, we did not want to run the risk of using a robot that was not sufficiently anthropomorphic (Perani *et al.*, 2001; Tai *et al.*, 2004). Furthermore, this and similar robots have 'presence' that generally cannot be elicited by computer-animated artificial agents (Sanchez-Vives and Slater, 2005).[2] Finally, by using a state of the art robot, we can engage more productively with social robotics, a rapidly developing field (Dautenhahn, 2007; Kahn *et al.*, 2007). As artificial agents become part of our lives, appearing in a variety of domains from Hollywood movies and video games, through to clinical and educational settings (Aitkenhead and McDonald, 2006; Coradeschi *et al.*, 2006), research on how humans respond to such agents is increasingly important (Saygin *et al.*, 2010).

One key issue is what artificial agents should look like (MacDorman and Ishiguro, 2006; Seyama and Nagayama, 2007; Kanda *et al.*, 2008). There is a wide range in what people may consider as an animate agent, as exemplified by well-known robots from cinema: from *HAL's* single camera eye, *R2D2, Wall-E* and *Eva*, which become surprisingly expressive and likeable with simple but effective designs, to more and more humanoid appearances such as the *Terminator, Robocop* and the replicants of *Blade Runner*.

It may seem like a good idea to make artificial agents look as human-like as possible, especially if they will be used in social settings. However, we soon encounter the 'uncanny valley': as an agent's appearance is made more human-like, people's disposition toward it becomes more positive, until a point at which increasing human-likeness leads to the agent being considered strange, unfamiliar and disconcerting. This phenomenon was prominently described in robotics (Mori, 1970), although there are early 20th century references to related concepts ['unheimlich', Freud, 1919; Jentsch, 1995 (1906)]. More recently, the uncanny valley has increasingly been experienced by the public when characters in movies or video games appear to be 'not quite right'. For example, many viewers found characters in the animated film *Polar Express* to be off-putting (Levi, 2004). Most modern androids, including Repliee Q2 used here, are also thought to fall into the uncanny valley (Ishiguro, 2006). Although the uncanny valley remains an influential concept due to substantial anecdotal evidence, and its importance for the design of artificial agents, there has been little systematic exploration of the phenomenon or its neural basis (Seyama and Nagayama, 2007; MacDorman *et al.*, 2009a; Steckenfinger and Ghazanfar, 2009).

Here, we hypothesized that the uncanny valley may, at least partially, be caused by the violation of the brain's

[1]The word android originates from a Greek root meaning 'man'. This is a gender-specific root, but in present day English the usage is generally gender neutral. When possible, we promote the use of 'humanoid' to refer to artificial agents modeled after humans, but this word does not allow us to distinguish between our experimental conditions in the present article.

[2]The scanner setup only allowed us to show a video of the robots to the subjects. Thus, while our setup did not allow for full presence, we studied robots that have presence in their normal setting.

predictions: When an agent looks like a human, based on a lifetime of experience, the brain generates a prediction that this appearance will be associated with a particular kind of behavior (e.g. movement kinematics). When the behavior of the agent violates the prediction, an error is generated (see 'Discussion' section; Rao and Ballard, 1999; Friston, 2010); although to be clear, prediction error is not the same thing as consciously experienced surprise (Friston, 2005; Kiebel *et al.*, 2009).

A related computational framework is provided by work on internal models of motor control (Wolpert *et al.*, 1995). When we perform an action, we predict the sensory consequences of that action through generative or forward models (Wolpert *et al.*, 1995; Wolpert and Miall, 1996). These predictions can be used to correct for unanticipated events, and to account for sensory noise and delays. The models can be recruited to infer the meaning of a perceived action given the sensory information (Wolpert *et al.*, 2003). During perception, the error between the prediction coming from internal models and incoming visual information can be minimized by selecting models yielding accurate predictions, that therefore correspond to the observed action (Kilner *et al.*, 2007).

To summarize, we performed functional magnetic resonance imaging (fMRI) as participants viewed short video clips of human or robotic agents carrying out recognizable actions. To our knowledge, the present study is the first neuroimaging investigation of action observation that has used robots with different levels of humanoid appearance. We used the android Repliee Q2, which has a very human-like appearance. With brief exposures, Repliee Q2 can be mistaken for a human being, but existing evidence indicates an uncanny valley experience with more prolonged exposure (Ishiguro, 2006). Importantly, we showed clips of Repliee Q2 both with its full human-like appearance, and also with a mechanical appearance, after stripping the robot of its human-like form, but retaining exactly the same mechanical movements. We also showed clips of the real human that Repliee Q2 was designed to replicate in appearance (Figure 1). There were thus three Agent conditions: Human, Android and Robot, which relate to our experimental interests of *appearance* and *motion* as follows: human and Android conditions feature biological (i.e. human-like) surface appearance, whereas the Robot condition features a mechanical appearance. In terms of motion, the Android and Robot feature nonhuman motion, whereas biological motion is unique to the Human condition. In this scheme, the Robot and the Human are different from each other in both dimensions, while sharing a feature with the Android. But from another perspective, the Robot and the Human conditions are similar in that they both feature congruent appearance and motion (looks human, moves human; looks mechanical, moves mechanical) whereas the Android features mismatching or incongruent appearance and motion (looks human, moves mechanical).

One limitation for most fMRI studies on this topic to date is that they compared the overall level of BOLD signal across conditions. fMRI can be used to allow more refined inferences regarding the neural representations underlying the measured activity. A well-established approach involves repetition suppression (also called fMRI adaptation): this method has its origins in neurophysiology, and refers to the phenomena of reduced neural response to a repeated stimulus (Henson and Rugg, 2003; Grill-Spector *et al.*, 2006; Krekelberg *et al.*, 2006). Repetition is thought to lead to such reduced responses only in neurons selective for the repeated properties, which allows the technique to be used as a means to explore what is represented in a particular brain region (e.g. motion direction sensitivity in area MT/V5 (Bartels *et al.*, 2008)). Repetition suppression effects are thought to reflect stimulus processing rather than task demands (Xu *et al.*, 2007) and observed attentional modulations are not generic (Thompson and Duncan, 2009). Recently, the repetition suppression approach has begun to be applied to the study of action perception to identify functional properties of the APS (e.g. Hamilton and Grafton, 2006, 2008; Dinstein *et al.*, 2007; Chong *et al.*, 2008; Fujii *et al.*, 2008; Lestou *et al.*, 2008; Kilner *et al.*, 2009).

The repetition suppression approach was ideally suited to our goals. BOLD differences for the experimental conditions (e.g. a main effect of Agent) can arise due to a number of low-level stimulus factors such as differences in illumination,

Fig. 1 Still images from the videos used in the experiment, depicting the agents. (**A**) Robot, (**B**) Android and (**C**) Human.

spatial frequency, color, or contrast that have little or nothing to do with action processing, as well as nonspecific attention or arousal effects. Instead, we focused on the interaction between repetition suppression and our experimental conditions.

METHODS

Participants

Twenty healthy adults (aged 20–36 years) participated. Data from one participant could not be used due to excessive head movement. All participants had normal or corrected vision, no cognitive, attentional or neurological abnormalities by self-report, and were right-handed. All participants gave written informed consent in accordance with local ethics approval.

Stimuli

Stimuli were video clips of actions performed by Repliee Q2 (in Android or Robot appearance, Figure 1A and B) and by the human 'master', after whom Repliee Q2 was modeled (Figure 1C). We refer to these agents as the Android, Robot and Human conditions (even though the former two are in fact the same robot).

Repliee Q2 has 42 degrees of freedom and can make head and upper body movements. In its existing implementation, it is impossible for this machine to exactly match the dynamics of human body movement (Pollick *et al.*, 2005). The actuators for Repliee Q2 were programmed over several weeks at Osaka University. The same movements were videotaped in two appearance conditions. For the Robot condition, we removed as many of the surface elements of Repliee Q2 as possible to reveal the materials underneath (e.g. wiring, metal arms and joints). The silicone 'skin' on the hands and face and some of the fine hair around the face could not be removed and was covered. In the Robot condition, Repliee Q2 could no longer be mistaken for a human (Figure 1A).

Crucially, the kinematics of the movement for the Android and Robot conditions were identical, since these conditions in fact comprised the same robot, carrying out the very same, programmed movements.

For the Human condition, we videotaped the female adult whose face was molded and used in constructing Repliee Q2. She was asked to watch each of the 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 eight actions per actor were used in the fMRI 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 (to convey no) and introducing self (Japanese bow)]. Video recordings were digitized, converted to grayscale and cropped to 400×400 pixels. A semi-transparent white fixation cross (40 pixels across) was superimposed at the center of the movies.

Experimental procedures and data analysis

MATLAB (Mathworks, Natick, MA, USA) and the Cogent toolbox (www.vislab.ucl.ac.uk/Cogent) were used for stimulus presentation and response collection.

Each participant was given exactly the same introduction to the study and the same exposure to the videos prior to scanning, because prior knowledge can affect judgments of artificial agents differentially (Saygin and Cicekli, 2002). To minimize possible effects of familiarity or expertise on our results, we only recruited participants who had no experience working with robots, had not spent time in Japan, nor had close friends or family from Japan (MacDorman *et al.*, 2009b). At the start of the study, subjects viewed each movie once outside of the scanner, and were told whether each agent was a human or a robot. They were not uncertain about the identity of the android by the time scanning took place.

Each participant was scanned in 6 445-s runs of the experiment, each comprising 12 blocks. Each block contained 12 videos from Human, Android or Robot conditions, presented in blocked counterbalanced order. Repetitions were event related. Videos were 2 s long and were separated by 500 ms. Each clip was equiprobably a repeat of the previous clip or a nonrepeat. Repetition intervals were kept constant between the conditions. Repetition suppression was calculated as the difference between BOLD response to a new (nonrepeated) stimuli compared with the response to the same stimulus when it was repeated. Positive suppression means there was less response to repeated stimuli. Additional illustrations of the experiment are shown in Supplementary Figure S1.

To ensure sustained attention, every 30 s, participants were presented with a written statement about which they had to make a True/False judgment (e.g. 'I did not see her waving') using an MRI compatible keypad. Participants had a maximum of 4 s to respond to each statement. We explored with a repeated measures analysis of variance (ANOVA) whether accuracy varied across conditions (it did not). Participants were instructed to keep their eyes on the fixation cross as much as possible, except at the end of the blocks when they read the statements. We used an MR-compatible eye tracker (see Supplementary Data) to check whether eye movements differed between conditions (they did not).

We used a 3 T Siemens Allegra scanner and a standard gradient echo pulse sequence. fMRI data were analysed with SPM 5 (http://www.fil.ion.ucl.ac.uk/spm) using standard procedures (see Supplementary Data for details). Although there is no agreed-upon localizer for the APS (Grafton and Hamilton, 2007), we selected regions of interest (ROIs), while also avoiding nonindependence errors (Kriegeskorte *et al.*, 2009) using the main effect of Repetition.

Table 1 Repetition suppression results from the whole brain random effects analysis

Anatomical description	BA	Peak (MNI)			*Z*	Mean RS (% Signal)			Agent differences
		x	*y*	*z*		Robot	Android	Human	
Temporal cortex									
Lateral temporal cortex (EBA)	37, 22	−48	−72	6	7.47	0.51	1.20	1.08	**Agent × repetition (*P* = 0.03)** *H > R (P = 0.07)* A > R (*P* = 0.02)
		50	−64	0	6.94	0.85	1.07	0.95	None
Fusiform gyrus		46	−44	−16	4.02	0.22	0.87	0.41	*Agent × repetition (P = 0.075)* A > R (*P* = 0.01) A > H (*P* = 0.05)
		−44	−44	−18	3.54	0.21	0.69	0.42	*A > R (P = 0.06)*
Occipital cortex									
V1/V2	17, 18	16	−88	2	5.86	−0.79	−0.88	−0.73	None
		−12	−94	2	5.73	−0.81	−0.87	−0.85	None
Parietal cortex									
sIPS	7	18	−72	60	4.93	0.29	0.88	0.51	A > R (*P* = 0.03)
		−20	−70	60	4.57	0.23	1.06	0.55	**Agent × repetition (*P* = 0.04)** A > R (*P* = 0.01) *A > H (P = 0.1)*
aIPS	40	−42	−38	42	3.96	0.30	0.81	0.42	**Agent × repetition (*P* = 0.002)** A > R (*P* = 0.002) A > H (*P* = 0.004)
		42	−36	42	4.37	0.22	0.93	0.39	**Agent × repetition (*P* = 0.002)** A > R (*P* = 0.003) A > H (*P* = 0.02)
Cuneus (pIPS)	19	24	−84	44	3.51	0.41	0.38	0.48	None
		−28	−82	40	3.63	0.41	0.41	0.39	None
Frontal cortex									
Middle Frontal Gyrus	10	−44	52	12	4.12	0.66	0.57	0.27	None
	10	46	50	−12	4.10	0.46	0.41	0.55	None
	46	50	48	10	3.93	0.58	0.80	0.22	*A > H (P = 0.07)*
	6	44	8	54	3.91	0.27	0.58	0.64	None
Other									
Parahippocampal/Amygdala		26	−4	−18	3.99	0.10	0.68	0.75	*A > R (P = 0.09)* *H > R (P = 0.08)*
		−28	−6	−20	3.69	0.20	0.58	0.66	None
Temporoparietal junction (TPJ)	40	60	−40	26	3.83	0.29	0.54	0.57	None
	40, 13	−48	−38	28	3.43	0.54	0.44	0.27	None
Cerebellum		8	−44	−18	3.69	0.48	0.49	0.46	None
Paracentral	5	4	−36	70	3.64	0.19	0.52	0.50	None
Postcentral gyrus	3	70	−8	24	3.45	0.41	0.45	0.35	None

Anatomical description and Brodmann Areas (BA) and the peak MNI coordinates are reported for each region in which the main effect of RS was significant ($P < 0.05$, FDR corrected and minimum cluster size of 30 voxels). Mean repetition suppression (percentage of signal change for Nonrepeat–Repeat, see 'Methods' section) for the three agents at these peaks are reported, along with any significant statistical differences (as measured using repeated measures ANOVA). We also noted pair-wise agent differences that were significant ($P < 0.05$ corrected, two tailed), and in italics, those that fell short of significance but with a tendency ($P < 0.1$, corrected, two tailed, denoted in italics). Significant Agent by Repetition interactions are marked in bold, and are also plotted in Figure 3.

Focusing on brain areas that are sensitive to action repetition, we explored contrasts of interest (Agent by Repetition interaction). We identified regions showing repetition suppression (Nonrepeat > Repeat) at $t \geq 8.86$; $P < 0.05$ false discovery rate (FDR) corrected, with a cluster size of at least 30 voxels (Table 1), and extracted percent signal change within a sphere of 5 mm radius around these peaks for each condition from each subject's first level analysis, and tested the Agent by Repetition interaction with an ANOVA. In a balanced factorial design with equiprobable conditions (as was used here), this process does not bias the chances of finding an interaction (Kriegeskorte *et al.*, 2009).

In reporting the effects, *P*-values were calculated two-tailed, and were corrected for multiple comparisons.

RESULTS

Behavioral and eye movement data

Mean Accuracy for the comprehension questions was 0.84 (s.d. = 0.28). Accuracy did not differ between conditions ($P > 0.1$ for all pair-wise comparisons). None of the eye movement measures (Mean and s.d. of *x* and *y position*, Pupil size) differed between conditions ($P > 0.1$ for all pair-wise comparisons). These data indicate comparisons across

conditions that were not subject to gross attention or eye movement confounds.

FMRI data

There were notable differences in repetition suppression between the agents, with the Human and Robot conditions leading to similar patterns of suppression, but the Android condition being distinctive, and leading to repetition suppression in a wider network (Figure 2). All agent conditions revealed repetition suppression in lateral temporal cortex. For the Android condition, repetition suppression was also evident in additional regions, notably in parietal and frontal cortex.

To confirm and quantify these results, we performed ROI analyses. Broadly consistent with previous repetition suppression studies of action perception, the main effect of Repetition revealed a network of areas, including occipital, lateral and ventral temporal, parietal, frontal, parahippocampal and cerebellar regions (Figure 3 and Table 1). All showed reduced responses to the repeated stimuli, with the exception of primary visual cortex, which showed repetition enhancement. Repetition suppression was found in the parietal and temporal nodes of the APS, but despite being a key node of the APS, ventral premotor cortex did not show significant repetition suppression (cf Chong *et al.*, 2008; Lestou *et al.*, 2008; Grossman *et al.*, 2010). There were other repetition suppression foci in frontal cortex, including one that extended into dorsal premotor cortex.

Since our main interest was differential responses to the three agents (and the stimulus dimensions they represent, i.e. biological appearance and motion), we tested whether the repeated measures ANOVA revealed a significant Agent by Repetition interaction. Even though there was qualitatively more suppression for the Android condition compared with the others in a widespread network (Figure 2), the Agent by Repetition interaction reached significance in only a subset of these regions (Table 1).

Figure 3 depicts repetition suppression as percent signal change in the peaks where the interaction was significant. In three parietal peaks, suppression was stronger for the Android condition than for the Human and Robot conditions: anterior intraparietal sulcus bilaterally (aIPS, Figure 3A and C), and a more posterior and superior parietal region (sIPS, Figure 3B) in the left hemisphere.

The Agent by Repetition interaction was also significant in left lateral temporal cortex, where we observed greater repetition suppression for the Human and Android conditions than for the Robot condition (Figure 3D). There was a large

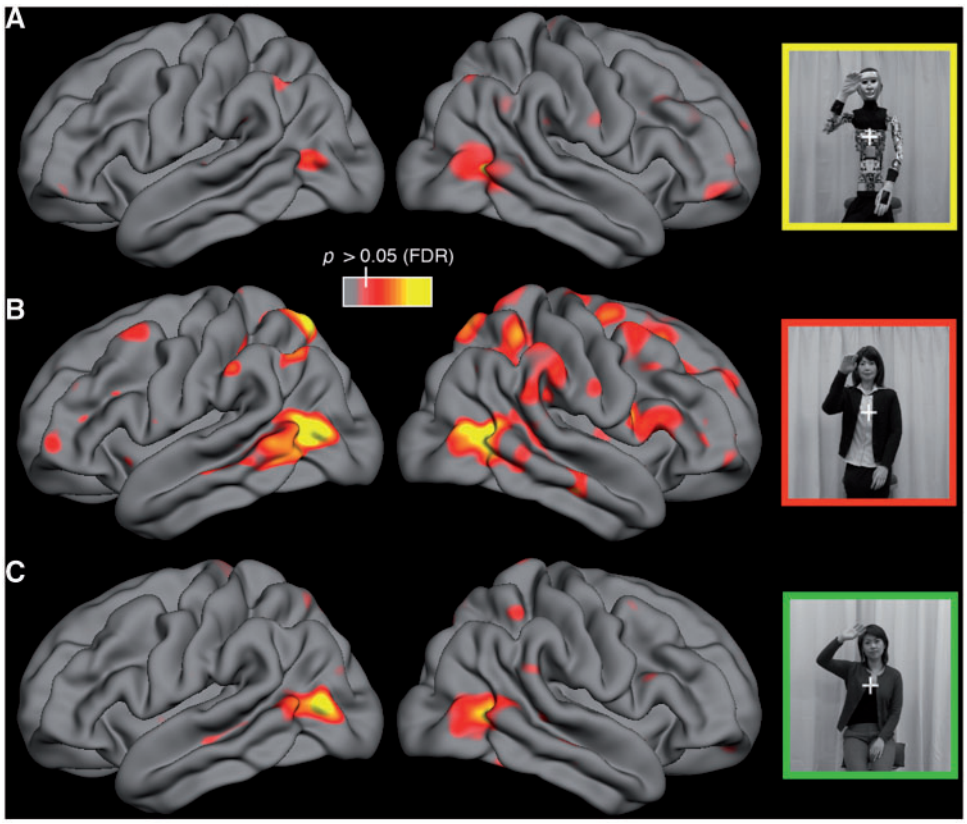

Fig. 2 Repetition suppression. Whole-brain repetition suppression effect for (**A**) Robot, (**B**) Android and (**C**) Human conditions rendered on the lateral views of the cortical surface of each hemisphere.

swathe of suppression covering multiple functional visual areas, but the interaction was present only in one subpeak, the coordinates of which corresponded to the previously reported location of the extrastriate body area (EBA) (Peelen and Downing, 2007), a region that responds more strongly to images of bodies and body parts compared with other kinds of stimuli. The right hemisphere peak, and a more dorsal subpeak of the left hemisphere cluster, did not show significant differences between agents.

For completeness, we also report the main effect of Agent: this effect was found in visual cortex bilaterally (with peaks in MNI coordinates −30, −92, 2 and 38, −80, −16), and was driven by a stronger response for the Robot condition compared with the other agents. These differences almost certainly reflect low-level visual differences between the stimuli (e.g. higher contrast, spatial frequency), demonstrating the advantage of using a repetition paradigm.

DISCUSSION

Summary of study and findings

We conducted this study as part of our general goal of identifying the functional properties of brain systems that allow us to understand others' body movements and actions (Saygin *et al.*, 2004b; Saygin, 2007). Subjects viewed actions performed by three agents that represented our experimental factors of interest: Human (biological motion and appearance), Android (biological appearance, nonbiological motion) and Robot (same agent as the Android, but 'skinned' to reveal the internal mechanics, nonbiological appearance and motion).

There was little evidence for specificity for biological motion or appearance *per se* in our data. Even though the nervous system processes form and motion in partially segregated systems, these attributes are inextricably interconnected (Shepard, 2001) and for action perception, the integration of motion and form cues may be a natural and critical aspect of the underlying computations.

There was a significant Agent by Repetition interaction in the anterior portion of the intraparietal sulcus bilaterally, corresponding to area aIPS, the putative human homologue of macaque area AIP (Grefkes and Fink, 2005; Culham and Valyear, 2006; Grafton and Hamilton, 2007). Here, suppression effects were larger for the Android compared with both the Human and the Robot conditions (Figure 3).

We found one region in left posterior lateral temporal cortex, where suppression for the Robot condition was significantly less than that for the human and the android, the two agents with human-like surface appearance. The peak location in this cluster corresponded the EBA (Peelen and Downing, 2007), consistent with the role of form-based processing in action perception (e.g. Lange and Lappe, 2006).

Predictive coding

We suggest that our results, especially the distinctive effects for the Android condition, can be reconciled with the 'predictive coding' framework of neural processing (e.g. Rao and Ballard, 1999; Friston, 2005, 2010; Kilner *et al.*, 2007; Jakobs *et al.*, 2009), which is based on minimization of prediction error among the levels of a cortical hierarchy. The key idea in this context is that brain activity will be higher for a stimulus that is *not* well-predicted or explained by a generative neural model of the external causes for sensory states (Friston, 2010). Predictive coding fits well with the view of perception as an active process that involves generating predictions about the environment, as well as the brain's own states (e.g. Yuille and Kersten, 2006; Bar, 2009; Barsalou, 2009).

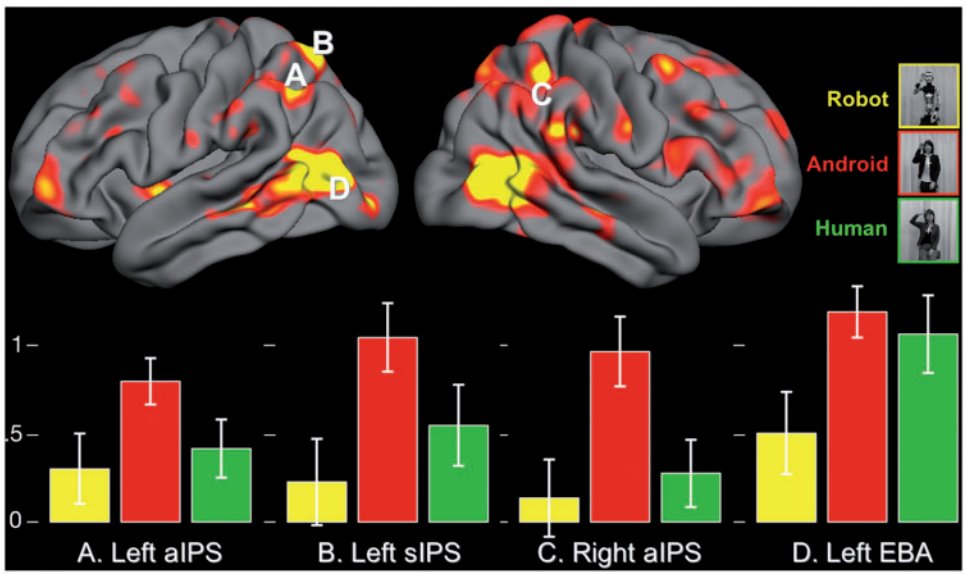

Fig. 3 Interactions. The top panel shows the main effect of Repetition (irrespective of Agent) rendered on the lateral views of the cortical hemispheres. The graphs depict the repetition suppression effect in all the peaks in which there was a significant interaction of Repetition by Agent (see Table 1 for statistics). *Y*-axes are percent signal change (Nonrepeat - Repeat).

We have a lifetime of experience that associates human appearance with biological motion, and machines (such as robots) with mechanical motion. For both our Human condition and our Robot condition, the observed motion kinematics was congruent with what would be predicted from the appearance of the agent. For the Android, however, there was a mismatch between the human-like appearance and the mechanical motion, leading to a larger prediction error, manifest as activity in relevant brain regions. A closer look at the data showed that responses to the nonrepeated videos were significantly greater for the Android compared with the other agents (Supplementary Figure S2), further supporting this interpretation. The prediction error would be smaller when a stimulus was preceded by the same stimulus, consistent with neural models of repetition suppression (Desimone and Duncan, 1995; Friston, 2005; Grill-Spector *et al.*, 2006).

The differences between agent types for repetition effects were most pronounced in parietal cortex. The aIPS, being the anatomical link between the posterior, visual components of the APS and the anterior, motor components (Petrides and Pandya, 1988; Seltzer and Pandya, 1994; Matelli and Luppino, 2001; Rozzi *et al.*, 2006), is ideally located to generate sensory predictions in this network. To describe the flow of information in the system, more time-resolved measurement techniques should be used, such as electroencephalography (EEG) or magnetoencephalography (MEG), as we are doing in related work.

Predictive coding not only provides a satisfactory interpretation of the current data, but also couches them in a framework that has both established and growing support in neuroscience (Rao and Ballard, 1999; Friston, 2005; Bar, 2009). We speculate that the present results reflect relatively general principles of neural organization, but also that the prediction errors may be dependent on how narrowly tuned the nervous system is for a particular domain. Future work should explore whether the perception of our conspecifics is an especially narrowly tuned domain, based on its evolutionary importance, and/or our extensive experience of interacting with conspecifics.

Contribution to the understanding of the uncanny valley

The uncanny valley has many potential dimensions (MacDorman and Ishiguro, 2006; Ho and MacDorman, 2010; Pollick, 2009). Our experiments and similar studies (e.g. Steckenfinger and Ghazanfar, 2009) were not designed in an optimum fashion to 'explain' the uncanny valley and as such can only make a modest contribution to defining its neural basis. However, the present results suggest an intriguing link between brain responses in the APS and the uncanny valley. While the android used in our study is often mistaken for a human at first sight, longer exposure and dynamic viewing has been linked to the uncanny valley (Ishiguro, 2006). In a predictive coding account of action perception, the android is not predictable—an agent with that appearance (human) would typically not move mechanically. When the nervous system is presented with 'the thing that should not be' [Lovecraft, 1984 (1936); Hetfield *et al.*, 1986], a propagation of prediction error may occur in the APS. While we cannot state a conclusive or causal link between prediction error and the uncanny valley based on the present data, we suggest this framework may contribute to an explanation for the uncanny valley.

Toward an interdisciplinary science of social perception

Humanoid robots and artificial agents are increasingly part of our daily lives (Kanda *et al.*, 2004; Dautenhahn, 2007; Tapus *et al.*, 2007). With application in domains such as healthcare, education, communications, entertainment and the arts, exploring human factors in the design and development of artificial agents is ever more important. This will require an interdisciplinary approach, to which we have contributed new data from cognitive neuroscience.

The present study is only a beginning. Computational modeling, ideally in conjunction with neuroimaging, will be important to specify or constrain the mechanisms underlying action perception, and to link this work with established frameworks of sensorimotor control (Wolpert *et al.*, 1995, 2003; Kawato, 1999; Kilner *et al.*, 2007). Predictive coding can be used to specify new hypotheses to explore further the interplay between appearance and motion of artificial agents, and to extend the approach to sensory integration more broadly. For example, it is possible that we have some prior idea of how robots should move—perhaps as evidenced by professionals making money by painting themselves gold, standing in front of cathedrals and moving like robots—and similar patterns of prediction errors for viewed actions might be generated for humans moving like robots (cf Shimada, 2010) or more generally, for other kinds of expectation violations between appearance and motion. Alternatively, the effects observed here could be specific to the perception of animate, or biologically relevant entities. Computer animation will be used to manipulate appearance and movement more parametrically and address these and similar questions in future work.

Despite many unknowns, our results already suggest an interpretation for the classic anecdotal reports of the uncanny valley effect. Psychologists have long pointed out those aspects of our physical experience that shape our perceptual systems (Gibson, 1979; Barlow, 2001). It has also long been acknowledged that violating perceptual expectations can have striking effects, compellingly illustrated by perceptual illusions (e.g. Gregory, 1980). As human-like artificial agents become more commonplace, perhaps our perceptual systems will be retuned to accommodate these new social partners. Or perhaps, we will decide it is not a good idea to make them so closely in our image after all.

SUPPLEMENTARY DATA

Supplementary data are available at *SCAN* online.

Conflict of Interest

None declared.

REFERENCES

Aitkenhead, M.J., McDonald, A.J.S. (2006). The state of play in machine-environment interactions. *Artificial Intelligence Review*, *25*, 247–76.

Bar, M. (2009). Predictions: a universal principle in the operation of the human brain. Introduction. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*, 1181–2.

Barlow, H. (2001). The exploitation of regularities in the environment by the brain. *Behavioral and Brain Science*, *24*, 602–7.

Barsalou, L.W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*, 1281–9.

Bartels, A., Logothetis, N.K., Moutoussis, K. (2008). fMRI and its interpretations: an illustration on directional selectivity in area V5/MT. *Trends in Neurosciences*, *31*, 444–3.

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

Buccino, G., Lui, F., Canessa, N., et al. (2004). Neural circuits involved in the recognition of actions performed by nonconspecifics: an FMRI study. *Journal of Cognitive Neuroscience*, *16*, 114–26.

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*, 1905–10.

Candidi, M., Urgesi, C., Ionta, S., Aglioti, S.M. (2008). Virtual lesion of ventral premotor cortex impairs visual perception of biomechanically possible but not impossible actions. *Society of Neuroscience*, *3*, 388–400.

Catmur, C., Walsh, V., Heyes, C. (2007). Sensorimotor learning configures the human mirror system. *Current Biology*, *17*, 1527–31.

Chaminade, T., Hodgins, J.K. (2006). Artificial agents in social cognitive sciences. *Interaction Studies*, *7*, 347–53.

Chaminade, T., Hodgins, J., Kawato, M. (2007). Anthropomorphism influences perception of computer-animated characters' actions. *Social Cognitive and Affective Neuroscience*, *2*, 206–16.

Chaminade, T., Zecca, M., Blakemore, S.-J., et al. (2010). Brain response to a humanoid robot in areas implicated in the perception of human emotional gestures. *PLoS ONE*, *5*(7), e11577.

Chong, T.T., Cunnington, R., Williams, M.A., Kanwisher, N., Mattingley, J.B. (2008). fMRI adaptation reveals mirror neurons in human inferior parietal cortex. *Current Biology*, *18*, 1576–80.

Coradeschi, S., Ishiguro, H., Asada, M., et al. (2006). Human-inspired robots. *IEEE Intelligent Systems*, *21*, 74–85.

Craighero, L., Metta, G., Sandini, G., Fadiga, L. (2007). The mirror-neuron system: data and models. *Progress in Brain Research*, *164*, 39–59.

Cross, E.S., Hamilton, A.F., Grafton, S.T. (2006). Building a motor simulation de novo: observation of dance by dancers. *Neuroimage*, *31*, 1257–67.

Culham, J.C., Valyear, K.F. (2006). Human parietal cortex in action. *Current Opinion in Neurobiology*, *16*, 205–12.

Dautenhahn, K. (2007). Socially intelligent robots: dimensions of human-robot interaction. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *362*, 679–704.

Desimone, R., Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.

Dinstein, I., Hasson, U., Rubin, N., Heeger, D.J. (2007). Brain areas selective for both observed and executed movements. *Journal of Neurophysiology*, *98*, 1415–27.

Freud, S. (1919). Das Unheimliche. SE, 17, 217–256. Translated at http://www-rohan.sdsu.edu/~amtower/uncanny.html (January 2011, date last accessed).

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *360*, 815–836.

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

Fujii, N., Hihara, S., Iriki, A. (2008). Social cognition in premotor and parietal cortex. *Social Neuroscience*, *3*, 250–60.

Gazzola, V., Rizzolatti, G., Wicker, B., Keysers, C. (2007). The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *Neuroimage*, *35*, 1674–84.

Gibson, J.J. (1979). *The ecological approach to visual perception*. Boston: Houghton- Mifflin.

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

Grefkes, C., Fink, G.R. (2005). The functional organization of the intraparietal sulcus in humans and monkeys. *Journal of Anatomy*, *207*, 3–17.

Gregory, R.L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *290*, 181–97.

Grill-Spector, K., Henson, R., Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*, 14–23.

Grossman, E.D., Jardine, N.A., Pyles, J.A. (2010). fMRI-adaptation reveals invariant coding for biological motion on the human STS. *Frontiers in Human Neuroscience*, *4*(15), 1–18.

Hamilton, A.F., Grafton, S.T. (2006). Goal representation in human anterior intraparietal sulcus. *Journal of Neurosciences*, *26*, 1133–7.

Hamilton, A.F., Grafton, S.T. (2008). Action outcomes are represented in human inferior frontoparietal cortex. *Cerebral Cortex*, *18*, 1160–8.

Ho, C.-C., MacDorman, K.F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, *26*(6), 1508–18.

Henson, R.N., Rugg, M.D. (2003). Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia*, *41*, 263–70.

Hetfield, J., Ulrich, L. Hammett, K. (1986). The Thing That Should Not Be. *Master of Puppets*, Electra Records. 12 inch Vinyl.

Ishiguro, H. (2006). Android science: conscious and subconscious recognition. *Connection Science*, *18*, 319–32.

Jakobs, O., Wang, L.E., Dafotakis, M., Grefkes, C., Zilles, K., Eickhoff, S.B. (2009). Effects of timing and movement uncertainty implicate the temporo-parietal junction in the prediction of forthcoming motor actions. *Neuroimage*, *47*, 667–77.

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*, 7315–29.

Jentsch, E. (1995 (1906)). On the psychology of the uncanny. *Angelaki*, *2*, 7–16.

Kahn, P.H., Ishiguro, H., Friedman, B., et al. (2007). What is a human? Toward psychological benchmarks in the field of human-robot interaction. *Interaction Studies*, *8*, 363–90.

Kanda, T., Ishiguro, H., Imai, M., Ono, T. (2004). Development and evaluation of interactive humanoid robots. *Proceedings of the IEEE*, *92*, 1839–50.

Kanda, T., Miyashita, T., Osada, T., Haikawa, Y., Ishiguro, H. (2008). Analysis of humanoid appearances in human-robot interaction. *IEEE Transactions on Robotics*, *24*, 725–35.

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, *9*, 718–27.

Kiebel, S.J., Daunizeau, J., Friston, K.J. (2009). Perception and hierarchical dynamics. *Frontiers in Neuroinformatics*, *3*, 20.

Kilner, J.M., Friston, K.J., Frith, C.D. (2007). The mirror-neuron system: a Bayesian perspective. *Neuroreport*, *18*, 619–23.

Kilner, J.M., Neal, A., Weiskopf, N., Friston, K.J., Frith, C.D. (2009). Evidence of mirror neurons in human inferior frontal gyrus. *Journal of Neuroscience*, *29*, 10153–9.

Kilner, J.M., Paulignan, Y., Blakemore, S.J. (2003). An interference effect of observed biological movement on action. *Current Biology*, *13*, 522–5.

Krekelberg, B., Boynton, G.M., van Wezel, R.J. (2006). Adaptation: from single cells to BOLD signals. *Trends in Neurosciences*, *29*, 250–6.

Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S., Baker, C.I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience*, *12*, 535–40.

Lange, J., Lappe, M. (2006). A model of biological motion perception from configural form cues. *Journal of Neuroscience*, *26(11)*, 2894–2906.

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*, 324–41.

Levi, S. (2004). Why Tom Hanks is less than human: While sensors cannot capture how humans act, humans can give life to digital characters. *Newsweek*, *650*, 305–6.

Lovecraft, H.P. (1984 (1936)). The Shadow Over Innsmouth. In: Joshi, S.T., editor. *The Dunwich Horror and Others*. Sauk City, WI: Arkham House.

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

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

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

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

Mori, M. (1970). The uncanny valley. *Energy*, *7*, 33–5.

Mukamel, R., Ekstrom, A., Kaplan, J., Iacoboni, M., Fried, I. (2010). Single neuron responses in humans during execution and observation of actions. *Current Biology*, *20*, 750–6.

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*, 2194–203.

Peelen, M.V., Downing, P.E. (2007). The neural basis of visual body perception. *Nature Reviews Neuroscience*, *8*, 636–48.

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 Neurosciences*, *23*, 6819–25.

Perani, D., Fazio, F., Borghese, N.A., et al. (2001). Different brain correlates for watching real and virtual hand actions. *Neuroimage*, *14*, 749–58.

Petrides, M., Pandya, D.N. (1988). Association fiber pathways to the frontal cortex from the superior temporal region in the rhesus monkey. *The Journal of Comparative Neurology*, *273*, 52–66.

Pobric, G., Hamilton, A.F. (2006). Action understanding requires the left inferior frontal cortex. *Current Biology*, *16*, 524–9.

Pollick, F.E. (2009). In search of the Uncanny Valley. In: Daras, P., Ibarra, O.M., editors. *UC Media 2009*. Venice, Italy: Springer, pp. 69–78.

Pollick, F.E., Hale, J.G., Tzoneva-Hadjigeorgieva, M. (2005). Perception of humanoid movement. *International Journal of Humanoid Robotics*, *3*, 277–300.

Press, C., Gillmeister, H., Heyes, C. (2007). Sensorimotor experience enhances automatic imitation of robotic action. *Proceedings: Biological Sciences / The Royal Society*, *274*, 2509–14.

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 Review of Neuroscience*, *27*, 169–92.

Rizzolatti, G., Fogassi, L., Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Review Neuroscience*, *2*, 661–70.

Rozzi, S., Calzavara, R., Belmalih, A., et al. (2006). Cortical connections of the inferior parietal cortical convexity of the macaque monkey. *Cerebral Cortex*, *16*, 1389–417.

Sanchez-Vives, M.V., Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, *6*, 332–9.

Saygin, A.P. (2007). Superior temporal and premotor brain areas necessary for biological motion perception. *Brain*, *130*, 2452–61.

Saygin, A.P., Cicekli, I. (2002). Pragmatics in human-computer conversations. *Journal of Pragmatics*, *34*, 227–58.

Saygin, A.P., Chaminade, T., Ishiguro, H. (2010). The perception of humans and robots: Uncanny hills in parietal cortex. *Annual Meeting of the Cognitive Science Society*, August 2010, Portland, Oregon.

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

Saygin, A.P., Wilson, S.M., Hagler, D.J. Jr, Bates, E., Sereno, M.I. (2004b). Point-light biological motion perception activates human premotor cortex. *Journal of Neuroscience*, *24*, 6181–8.

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. *Journal of Comparative Neurology*, *343*, 445–63.

Seyama, J., Nagayama, R. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and Virtual Environments*, *16*, 337–51.

Shepard, R.N. (2001). Perceptual-cognitive universals as reflections of the world. *Behavior and Brain Science*, *24*, 581–601.

Shimada, S. (2010). Deactivation in the sensorimotor area during observation of a human agent performing robotic actions. *Brain and Cognition*, *72*, 394–9.

Steckenfinger, S.A., Ghazanfar, A.A. (2009). Monkey visual behavior falls into the uncanny valley. *Proceedings of the National Academy of Science USA*, *106*, 18362–6.

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*, 117–20.

Tapus, A., Matarić, M.J., Scassellati, B. (2007). The grand challenges in socially assistive robotics. *IEEE Robotics and Automation Magazine*, *14*, 35–42.

Thompson, R., Duncan, J. (2009). Attentional modulation of stimulus representation in human fronto-parietal cortex. *Neuroimage*, *48*, 436–48.

Wolpert, D.M., Doya, K., Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *29*, 358:593–602.

Wolpert, D.M., Ghahramani, Z., Jordan, M.I. (1995). An internal model for sensorimotor integration. *Science*, *269*, 1880–2.

Wolpert, D.M., Miall, R.C. (1996). Forward models for physiological motor control. *Neural Networks*, *9*, 1265–79.

Xu, Y., Turk-Browne, N.B., Chun, M.M. (2007). Dissociating task performance from fMRI repetition attenuation in ventral visual cortex. *Journal of Neuroscience*, *27*, 5981–85.

Yuille, A., Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences*, *10*, 301–8.